A TREATISE ON LIGHT

A TREATISE ON LIGHT

BY

R. A. HOUSTOUN, M.A., Ph.D., D.Sc.
LECTURER ON PHYSICAL OPTICS IN THE UNIVERSITY OF GLASGOW

WITH 328 *DIAGRAMS*

LONGMANS, GREEN AND CO.
39 PATERNOSTER ROW, LONDON
FOURTH AVENUE & 30TH STREET, NEW YORK
BOMBAY, CALCUTTA, AND MADRAS
1915

PREFACE.

This book is intended for students who have been through a first year's physics course and who are proceeding further with the study of light. It differs from other books on light by a more systematic treatment, also by dealing with the full scope of the subject and including the results of recent investigations.

A good knowledge of elementary mathematics is assumed. The calculus is used, but I hope that the results obtained by its aid will be intelligible to those who cannot follow the intermediate steps, and in any case the greater part of the book is free from it.

Of the 328 figures more than 270 have been specially drawn for the book.

Mr. Chas. Cochrane, M.A., B.Sc., has read all the proof-sheets and has worked all the examples.

The book is the fruit of nine years' teaching and experimenting at Glasgow University, and I am indebted to Prof. Gray and the other members of the staff of the Natural Philosophy Department, and also to some of the advanced students for much stimulation and encouragement. Prof. Gray has also read part of the proofs.

R. A. HOUSTOUN.

CONTENTS.*

PART I.

Geometrical Optics.

CHAPTER I. FUNDAMENTAL IDEAS.

CHAPTER II. ELEMENTARY THEORY OF SPHERICAL MIRRORS AND LENSES.

CHAPTER III. THICK LENSES AND SYSTEMS OF LENSES.

CHAPTER IV. THE DEFECTS OF THE IMAGE.

* The numbers give pages.

CHAPTER V. ON DETERMINING THE CONSTANTS OF MIRRORS AND LENSES.

CHAPTER VI. OPTICAL INSTRUMENTS.

CHAPTER VII. THE SPECTROMETER AND THE DETERMINATION OF INDICES OF REFRACTION.

PART II.

PHYSICAL OPTICS.

CHAPTER VIII. THE VELOCITY OF LIGHT.

CHAPTER IX. INTERFERENCE.

CHAPTER X. DIFFRACTION.

PART I.

GEOMETRICAL OPTICS.

CHAPTER I.

FUNDAMENTAL IDEAS.

LIGHT travels in straight lines. This is taken for granted by everyone, for we always assume that a body exists in the direction of the rays of light which enter our eye from it. Also numerous illustrations of the rectilinear propagation of light occur in daily life. For example, the rays of the sun entering a darkened room through a chink in a shutter are seen to be straight; also if a shadow of a stick be cast by a candle flame, the flame, the top of the stick, and the shadow of the top of the stick are all in one straight line.

But if observations are made with great accuracy, it is found that the propagation of light is only approximately rectilinear. If, for example, light from a point source falls on a screen with a very narrow hole in it, say $\frac{1}{10}$ mm. diameter, after passing through the hole the rays bend into the shadow to such an extent that we can no longer speak of the rectilinear propagation of light. In this case diffraction is said to take place. Diffraction will be fully considered in a subsequent chapter; it can be ignored as far as concerns the great majority of optical phenomena.

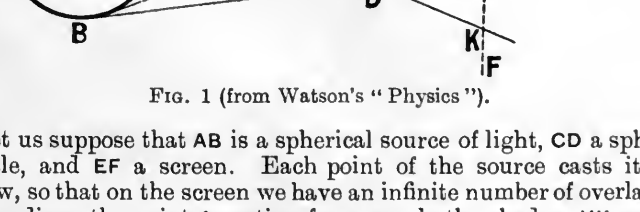


FIG. 1 (from Watson's "Physics").

Let us suppose that AB is a spherical source of light, CD a spherical obstacle, and EF a screen. Each point of the source casts its own shadow, so that on the screen we have an infinite number of overlapping shadow discs, the point A casting for example the shadow HK and the point B the shadow GI. As A and B are the extreme points of the source, the overlapping portion HI receives no light at all. The parts

GH and IK receive light from part of the sphere, and the shadow gradually becomes brighter as we proceed from H to G and from I to K. The whole shadow consists therefore of a perfectly black disc of diameter HI called the *umbra* surrounded by a ring of gradually diminishing darkness called the *penumbra*.

If the sphere AB represents the sun and the sphere CD the moon, it is only when a point on the earth's surface enters the cone CMD that the sun is totally eclipsed for that point. When it is within the penumbra, the eclipse is only partial.

Pinhole Camera. If a luminous object AB (fig. 2) is placed in front of a small hole OP in an opaque screen and a white card placed on the other side of the hole, an inverted image CD of the luminous object will be formed on the card. A pencil of light goes out from every point on the object in the manner indicated by the diagram, and forms a patch of light on the card. The hole must be small, literally a pinhole, otherwise the patches of light on the card will be too large and the image blurred. Consequently the image is very faint and if

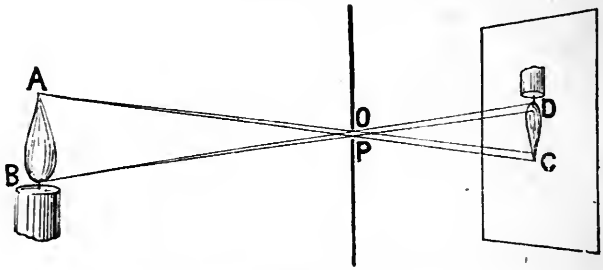


FIG. 2 (from Watson's "Physics").

a photograph is taken by this method a very long exposure is necessary. The pinhole camera has, however, the advantages that it takes in a very wide angle of view and also that no focussing is necessary, i.e. there is perfect depth of focus.

Laws of Reflection and Refraction. The fact that a body not itself luminous is yet visible in all directions, when illuminated by light from a light source, shows that it must be capable of reflecting light in all directions. Such reflection is called diffuse reflection and is subject to no simple law. It takes place at all rough surfaces. Rays that undergo diffuse reflection usually change their colour.

When a narrow pencil of light is reflected from a mirror or the polished surface of a transparent medium, diffuse reflection takes place to only a very small extent and the incident pencil of light gives rise to two pencils, a reflected one and a refracted one. The light is then said to undergo regular reflection. The point where the incident ray of light strikes the surface is called the point of incidence, and if a normal be drawn to the surface at this point, the angle which it makes with the incident ray is called the angle of incidence and the angle which it makes with the reflected ray is called the angle of reflection. If the normal be produced downwards into the transparent medium, the angle which it makes with the refracted ray is called the angle of refraction.

The laws of reflection and refraction are as follows:—

The incident ray, the normal to the reflecting surface at the point of incidence and the reflected ray are all in the same plane.

The angle of reflection is equal to the angle of incidence.

The refracted ray lies in the same plane as the normal and the incident ray, and is on the opposite side of the normal from the incident ray.

The sine of the angle of incidence bears a constant ratio to the sine of the angle of refraction for all angles of incidence, the value of the ratio depending on the nature of the light and on the nature of the media in contact at the surface at which refraction takes place.

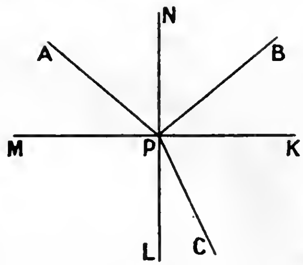


Fig. 3.

Thus in fig. 3 if P is the point of incidence, MK the trace of the reflecting surface, NP the normal and AP the incident ray, the reflected ray PB and the refracted ray PC are both in the same plane as AP and PN, $\angle$ NPB = $\angle$ NPA and sin NPA/sin LPC is constant, no matter what the value of $\angle$ NPA is.

Write
$$\frac{\sin \mathrm{NPA}}{\sin \mathrm{LPC}} = \mu.$$

Then μ is said to be the index of refraction of the lower medium with reference to the upper for the light in question. For all transparent solids and liquids μ is greater than 1.

When the angle of incidence is zero, AP and PB both coincide with NP, and PC coincides with PL. The ray is then undeviated by refraction.

If the surface on which the light falls is curved, we may divide it up into small elements of area, regard these elements as plane, draw a normal to each and consider each separately. The light falling on each will then be reflected and refracted according to the above laws.

The laws of reflection and refraction can best be proved with the spectrometer (p. 98), but a very simple means of verifying them in the case of a glass slab will be described here.

Fix a piece of paper on a drawing-board and draw any straight line AP on it. Place the slab in position so that one of its faces rests along MK and draw the line MK by running a pencil along its edge. Look into the surface of the slab and the line AP will be seen by reflection in the direction A′P; then place a ruler on the paper so that its edge seems a continuation of A′P and draw part of the line PB. Remove the slab and produce BP to cut MK. It should cut it at P, and it can be shown with a pair of compasses in the manner indicated in the diagram that ∠ APN = ∠ NPB.

Draw a straight line AP as before and place the slab in position so that MK and HJ are the traces of two of its parallel faces. Draw MK

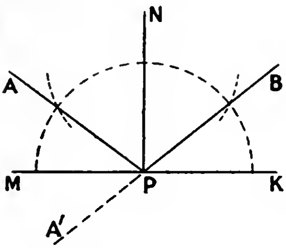


FIG. 4.

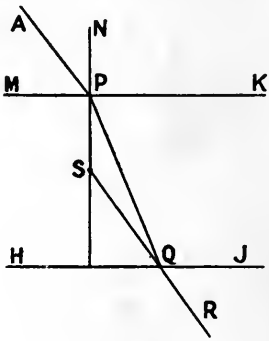


FIG. 5.

and HJ. Then looking into the face HJ of the slab draw with a ruler QR, which appears to be the continuation of the line AP. Remove the slab, draw the normal NPS, produce RQ to meet the normal at S and join PQ. Then ∠ SPQ is the angle of refraction.

The ratio of the sines can now be determined in different ways but perhaps the neatest is as follows. It is found from the drawing that SR is parallel to AP. Consequently ∠ PSQ is the supplement of ∠ APN and sin APN = sin PSQ. Hence, since the sines of the angles of a triangle are proportional to the lengths of the opposite sides,

$$\mu = \frac{\sin \mathrm{APN}}{\sin \mathrm{SPQ}} = \frac{\sin \mathrm{PSQ}}{\sin \mathrm{SPQ}} = \frac{\mathrm{PQ}}{\mathrm{SQ}}.$$

No matter how ∠ APN varies, the ratio PQ/SQ is always constant.

§ Let MN be the section of a plane mirror perpendicular to the plane of the paper, and let P be a point source of light. Let PA be any ray from P to the mirror; then after reflection it has the direction AC. Draw EA, the normal at A, draw PN perpendicular to MN and produce CA and PN to meet at Q.

In △s APN and AQN we have AN common, ∠ ANP = ∠ ANQ both being right and ∠ PAN = ∠ QAN, since ∠ EAN is right and EA bisects ∠ CAP. Hence the triangles are equal and PN = NQ.

The point Q is fixed and quite independent of the direction of PA. If we draw another ray PB, DB its direction after reflection must also pass through Q. All rays diverging from P and striking the mirror appear therefore to an eye at CD to diverge from Q. The point Q is thus said to be the image of P, and since the rays AC and BD themselves do not pass through Q but only their directions produced backwards do, the image is said to be virtual. If the rays themselves had passed through Q, the image would have been real. The essential difference between a real image and a virtual image is, that a real image can be received upon a screen and made visible; a virtual image cannot.

If instead of a luminous point we have a line object PR, an image is formed of every point of it as above, and these point images combine to form the line image QS as shown in fig. 7. It will be

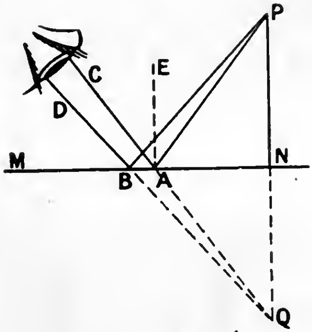


Fig. 6.

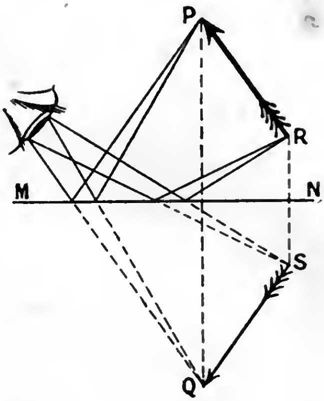


Fig. 7.

noticed that the object is reversed by reflection, that right and left are interchanged. Consequently, if a piece of writing be held up to the mirror, its image is illegible; if however it be blotted and the blotting-paper held up to the mirror, the writing, being reversed twice, becomes legible again.

The fact that the image is as far behind the mirror as the object is in front of it can be shown very easily with two pins, a short one and a long one, and a glass plate, the back of which has been blackened so that only the front surface can reflect light. The glass is mounted vertically and the short pin stuck up in front of it. The long pin is placed behind the glass in such a position, that it and the image of the short pin appear to occupy the same position. Then there is no parallax between the long pin and image of the short one, i.e. moving the eye from side to side does not cause the one to move past the other. If the adjustment is made accurately, the long pin and the short pin are equidistant from the surface of the glass.

If the experiment is performed with a piece of mirror glass instead of an unsilvered plate, the image is of course brighter, but a complication arises owing to the fact that the light has to pass through the glass before it gets to the mirror. It will be shown later that if t is the thickness of the glass, this makes the image nearer the mirror by $2(\mu - 1)t/\mu$. If the mirror is on the front surface of the glass, this complication falls away. But in this case the mirror has to be made by chemically depositing silver on the glass and is not very permanent; the ordinary tinfoil-mercury process can be applied only to the back of a glass surface.

Rotation of a Plane Mirror. Let MN be the intersection by the paper of a plane mirror perpendicular to the plane of the paper. Let PO be the incident ray, OQ the reflected ray, and OH the normal at the point of incidence. Let the mirror be rotated through an angle a about an axis through O perpendicular to the plane of the paper and let OR and OG be the positions of the reflected ray and normal after the rotation. Then

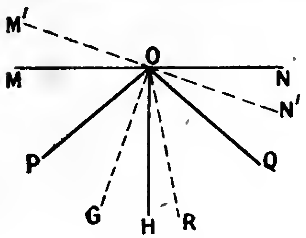


FIG. 8.

$$a = \angle M'OM = \angle GOH = \angle POH - \angle POG = \tfrac{1}{2}\angle POQ - \tfrac{1}{2}\angle POR$$
$$= \tfrac{1}{2}(\angle POQ - \angle POR) = \tfrac{1}{2}\angle QOR.$$

But QOR is the angle turned through by the reflected ray. Therefore the angle turned through by the reflected ray is twice the angle turned through by the mirror.

Use is often made of a mirror and reflected ray to measure the angle through which a body rotates. For example, in some patterns of galvanometer a small magnet or system of magnets has a little circular mirror attached, which reflects the light from a lamp on to a graduated scale usually at a distance of about a metre from the mirror. When a current passes, the magnet and mirror are deflected and consequently the image moves along the scale. The reflected rays thus form a perfectly straight mass-less pointer.

Multiple Reflections. ABCD is a thick plate of glass with parallel faces. A ray of light Pa falls on the upper surface of the plate at a and gives rise to a reflected ray aa' and a refracted ray aa''. The refracted ray falls on the lower surface of the plate and gives rise to the reflected ray $a''b$ and a refracted ray which is not shown. The reflected ray $a''b$ in turn falls on the upper surface of the plate at b and gives rise to the refracted ray bb' and the reflected ray bb''. And so on; the single incident ray Pa gives rise to an infinite number of rays aa', bb', cc', dd', etc., which appear to come from the images P_1, P_2, P_3, P_4, etc., to an eye above the plate. The first two, P_1 and P_2, are about equally bright and the others get rapidly fainter, because each reflection diminishes the intensity of the light very considerably. If the experiment is performed with a candle flame in a darkened room usually only three images can be seen. If, however, the lower surface of the glass is silvered, the second image is much brighter than the others and altogether about five images can be seen.

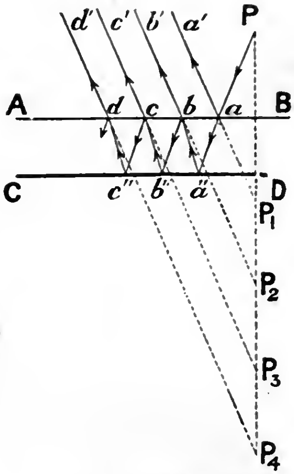


Fig. 9 (from Watson's "Physics").

In an ordinary mirror it is this second image that we see: the other images are there but are so faint that they can be ignored.

Suppose a point source P (fig. 10) is placed between two mirrors A and B which have their faces turned towards one another, so that its distance from A is a and its distance from B is b. Then by reflection in A an image is formed at A_1; by reflection in B an image of this image is formed at A_2, and by reflection again in A this in turn gives rise to an image at A_3, and so on. Similarly by considering the reflection of P in B we get B_1 and another infinite series of images, B_1, B_2, B_3 . . . There are thus two infinite series of images, but of course, owing to the light getting fainter by successive reflection it is only the first of each series that are seen. Above each image in the diagram is given its distance from P. The

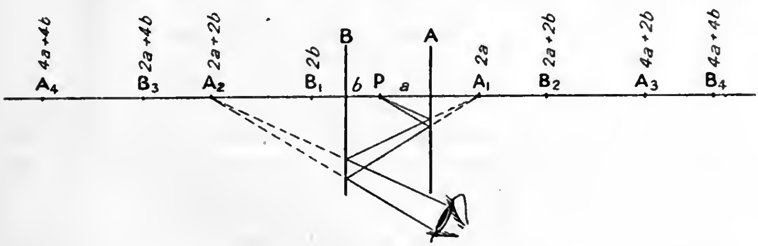


Fig. 10.

diagram also shows the path taken to the eye by the pencil of rays, which forms one of the images, A_2.

§ The positions of the images become interesting when the two mirrors are inclined to one another. In the figure AO and BO represent the mirrors and P is the object. P forms an image P_1 by reflection in OA. Since PP_1 is at right angles to OA and P_1 is as far behind OA as P is in front of it, P_1 is on the circle through P which has O as centre. The image P_1 forms an image P'' in OB. This image is obviously on the same circle as P and P_1. It in turn forms an image P_3 in OA. The line joining P'' and P_3 passes of course beyond the end of OA, but this is immaterial. For P'' to form an image in OA it is only necessary that rays from P'' should reach the eye after reflection at OA. P_3 forms an image P'''' in OB ; P'''' is behind both mirrors and can consequently produce no further images.

If we go back to the object P, it forms an image P' in OB which in turn

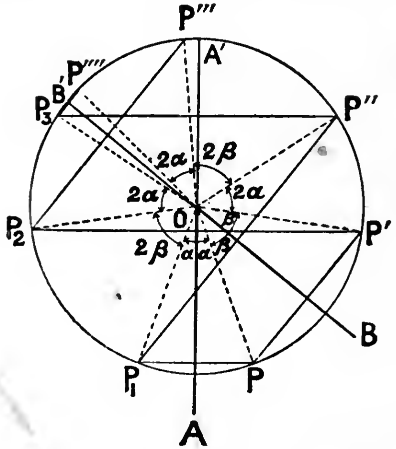


FIG. 11 (from Watson's "Physics").

forms an image P_2 in OA. The latter forms an image P''' in OB, and this image is behind both mirrors ; hence no further images are formed. Thus in the case in question seven images altogether are formed, and these all lie on the circle with centre O and radius OP.

The angles between the different images are marked on the figure, the angles AOP and BOP being respectively denoted by α and β, and it is found that the angle between two consecutive images is alternately equal to 2α and 2β.

Refraction through Slabs with Parallel Sides. Let a ray of light ABDE be refracted through a plate of a transparent material with parallel sides, as shown in fig. 12. Let μ_{12} be the index of refraction of the plate with reference to the medium in which it is placed, and let μ_{21} be the index of refraction of the medium with reference to the material of the plate. It is found by experiment that the direction of a ray is never altered by passing through a medium with parallel sides ; hence the emergent ray DE makes the same angle α with the normal as the incident ray does. If β denotes the angle of refraction,

$$\mu_{12} = \frac{\sin \alpha}{\sin \beta}, \qquad \mu_{21} = \frac{\sin \beta}{\sin \alpha}.$$

Therefore $\mu_{21} = 1/\mu_{12}$.

Suppose that two slabs with parallel faces are placed together and that a ray of light ABCDE passes through them as shown in fig. 13. Let μ_{12} be the index of refraction of the second medium with reference to the first, let μ_{23} be the index of refraction of the third medium with reference to the second, and let μ_{31} be the index of refraction of the first medium with respect to the third. Then, if β and γ have the values shown in the figure,

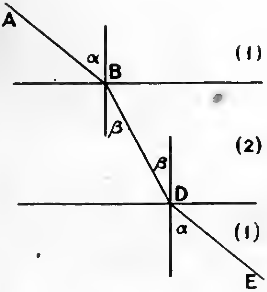


FIG. 12.

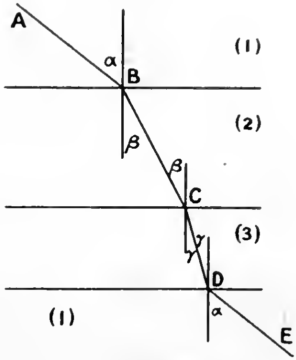


FIG. 13.

$$\mu_{12} = \frac{\sin \alpha}{\sin \beta}, \mu_{23} = \frac{\sin \beta}{\sin \gamma} \text{ and } \mu_{31} = \frac{\sin \gamma}{\sin \alpha} \quad . \quad . \quad (1)$$

Hence
$$\mu_{12}\,\mu_{23}\,\mu_{31} = \frac{\sin \alpha}{\sin \beta}\,\frac{\sin \beta}{\sin \gamma}\,\frac{\sin \gamma}{\sin \alpha} = 1,$$

and
$$\mu_{23} = \frac{1}{\mu_{12}\,\mu_{31}} = \frac{\mu_{13}}{\mu_{12}}.$$

The index of refraction with reference to a vacuum of air under standard pressure and at a temperature of 0° C. is 1·0003. The index of refraction with reference to a vacuum is usually called the absolute index of refraction, and, when the index of refraction of a gas is given, it is always its absolute index of refraction that is meant. When the index of refraction of a glass or a liquid is given, it is usually the index of refraction with reference to air that is meant. If either the index of refraction with reference to air or the absolute index of refraction is given, the other can easily be found by the equation $\mu_{23} = \mu_{13}/\mu_{12}$, since the absolute index of refraction of air is known.

If we denote the absolute indices of refraction of the three media in fig. 13 by μ_1, μ_2, μ_3, then $\mu_{12} = \mu_2/\mu_1$, $\mu_{23} = \mu_3/\mu_2$, and $\mu_{31} = \mu_1/\mu_3$; hence (1) can be written

$$\mu_1 \sin \alpha = \mu_2 \sin \beta = \mu_3 \sin \gamma.$$

This equation can easily be extended to take in the case of n media bounded by parallel plane surfaces, and we see from it that the inclination of the ray in any one medium depends only on the original inclination and is independent of the intermediate media passed through.

Astronomical Refraction. Since the index of refraction of air is appreciably greater than for a vacuum, when the rays from a star enter our atmosphere they are refracted. The effect of this refraction is to make the star appear higher in the heavens than it really is. Since the density and consequently the refractive index of the atmosphere decreases gradually as we ascend, the refraction does not take place all at once, but the rays are curved during their passage through the atmosphere. We can get the effect of this gradually decreasing index of refraction by supposing the atmosphere divided into parallel layers, for each of which the index of refraction is constant and has a smaller value than for the layer immediately below it.

Let fig. 14 represent the passage of a ray of light through such a layer. Then applying the law of refraction at the point P we have

$$\mu_1 \sin i = \mu_2 \sin r = \mu_2 \sin (i - \rho),$$

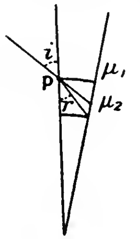


FIG. 14.

where ρ is the deviation produced by the refraction. Since ρ is small we can write $\sin \rho = \rho$, $\cos \rho = 1$, and the above equation becomes

$$\mu_1 \sin i = \mu_2 (\sin i - \rho \cos i)$$
$$\text{or } \rho = (\mu_2 - \mu_1) \tan i$$

since μ_2 can be written approximately equal to 1. If we form this equation in succession for all the other layers and add, and if we also assume that i is not near 90° and $\tan i$ consequently varies slowly, then the total deviation is given approximately by

$$\rho = (\mu - 1) \tan i,$$

where μ is the index of refraction of the atmosphere close to the earth's surface. The index of refraction in the upper region of the atmosphere is, of course, 1.

The full theory of astronomical refraction is very complicated, as the magnitude of the effect depends on the temperature and pressure of the atmosphere. The upward displacement produced increases very rapidly as the star approaches the horizon, attaining a magnitude of 35′ when i reaches 90°. This is greater than the angle subtended by the diameter of the sun or the moon. Thus when the lower edge of the sun's disc appears to be touching the horizon, the whole disc is really below the plane of the horizon. Hence owing to the refraction of the atmosphere both ends of the day are lengthened at the expense of the night.

§ It sometimes happens in a desert that the layer of air immediately above the sand is much hotter than the layers higher up, consequently its density and index of refraction are lower than for the layers higher up. Rays of light from the sky incident on the hot layer at a very large angle are thus totally reflected * and pass upwards again in the colder region without reaching the sand. If they reach the eye of an observer, he sees a piece of the sky apparently mirrored in the sand and takes it to be the surface of a lake. To this phenomenon the name of mirage is given.

If objects at a distance are regarded through a stream of hot air, e.g. through the hot air issuing from a chimney or rising from a seashore heated by the sun, they appear to wave about and tremble. This is due to the index of refraction of the hot air being less. The patches of hot air act like prisms, deviating the rays and changing the apparent directions from which they come and the positions and shapes of the patches of hot air are always altering. The scintillation or twinkling of the stars is similarly to be ascribed to changing inequalities in the refractive index of the atmosphere.

Image of a Point formed by Refraction at a Plane Surface. Let us suppose that a pencil of rays is diverging from a point P on the under surface of a parallel-sided slab of some refracting material, for example, glass. The ray PN which meets the upper surface of the slab at right

* For the meaning of this expression *cf.* next page.

angles does not change its direction in passing into the air, but any other ray PS is refracted at the surface of separation. Let SR be its direction after refraction and let μ be the index of refraction of the glass. Then

$$= \frac{\sin TSR}{\sin PSM} = \frac{\sin NQS}{\sin QPS} = \frac{\sin PQS}{\sin QPS} = \frac{SP}{SQ}.$$

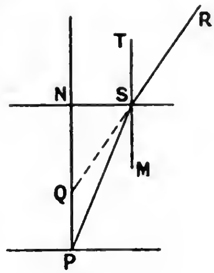


FIG. 15.

If the angle QPS is small, SP/SQ = NP/NQ : consequently μNQ = NP and the point Q is fixed. Thus, if a thin pencil from P be incident at right angles on the upper surface of the glass, it appears after refraction to come from Q, i.e. a virtual image of P is formed at Q.

An experimental method of finding the index of refraction of a glass slab is founded on this. A microscope is used, the tube of which can be raised vertically by a rack and pinion motion. This microscope is first focussed on a mark on the stage, then the slab of glass is placed on the mark and the microscope focussed on the virtual image of the mark formed by the slab. The distance through which the microscope has to be raised gives PQ. This is read on a scale. The microscope is then focussed on a scratch on the upper surface of the slab and the height through which it has to be raised this time gives QN. Then, when PQ and QN are known, the index of refraction can be calculated.

In determining the index of refraction of liquids by this method a scratch on the bottom of the vessel containing the liquid is taken as the first mark and dust or chalk floating on the surface of the liquid as the second mark. The method is not an accurate one and is used principally for illustrating the theory.*

If the rays diverge from P making a wide angle with PN, their directions after refraction no longer pass through Q but touch a curve as is shown in the figure. This curve is termed a caustic.

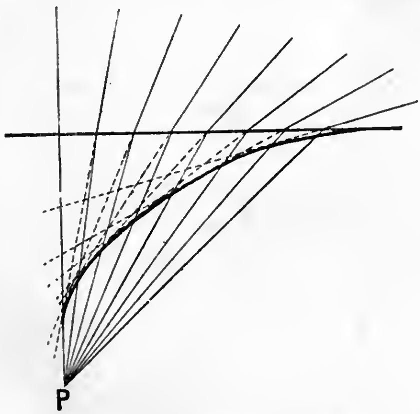


FIG. 16 (from Watson's "Physics").

Total Reflection. The angles of incidence and refraction are connected by the relation $\sin i = \mu \sin r$. If the ray is passing from a medium

* It can be improved by using a Ramsden eyepiece with cross-wires instead of the Huygens eyepiece usual with microscopes. Then the parallax between the cross-wires and image makes the focussing more accurate.

with a large index of refraction to one with a small index of refraction, say from glass to air, μ is less than 1. As i, the angle of incidence inside the glass increases, r increases, and when sin i becomes equal to μ, $\sin r = 1$ and $r = 90°$. The value of i given by $\sin i = \mu$ is known as the critical angle. When the angle of incidence is greater than the critical angle, sin r is greater than 1. Consequently r can have no real value; there is no refracted ray and the light is said to be totally reflected. In fig. 17 are shown corresponding angles of incidence and refraction for rays incident internally on the surface of glass of index of refraction 1·5. The ray PA makes the limiting angle with the normal. The ray PB is totally reflected.

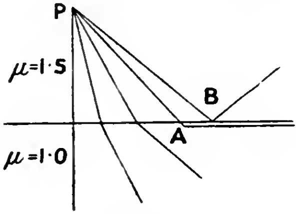


FIG. 17.

Aplanatic Surfaces. The optical distance between two points or the optical length of the path traversed by a ray between two points is equal to the actual distance multipled by the index of refraction of the medium containing the path. If the path goes through different media, then its optical length is obtained by multiplying the length of each part of it by the index of refraction of the medium, in which that part is, and then taking the sum.

An aplanatic surface is one for every point of which the sum of the optical distances from two fixed points is constant.

Let us suppose that the two fixed points S and S′ are in the same medium, that the light goes directly to the aplanatic surface and back, and that P is any point on the aplanatic surface. Then its equation is

$$SP + S'P = c,$$

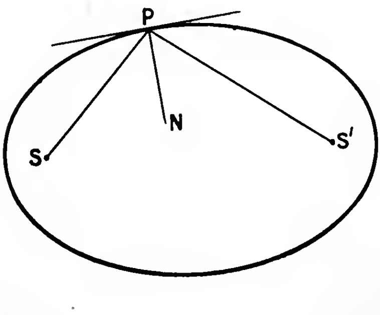


FIG. 18.

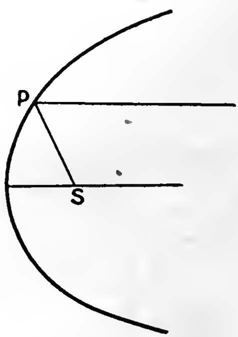


FIG. 19.

where c is a constant. This is the equation to an ellipsoid of revolution with S and S′ as foci, i.e. the surface obtained by rotating the ellipse in fig. 18 round the line SS′. Now a property of the ellipse is, that the straight lines joining any point on it to the foci make equal angles with the tangent and normal at that point. Consequently if PN is the normal at P, $\angle SPN = \angle S'PN$, and if the inner surface of the ellipsoid be regarded as a mirror, any ray of light incident on the surface from the one focus must pass after reflection through the other focus.

If one of the foci, S′, is moved to infinity, the ellipsoid becomes a paraboloid of revolution, and if S is regarded as a source of light, every ray that diverges from S, no matter at what angle, after reflection at the mirror becomes parallel to the axis. Conversely a paraboloidal mirror brings all the rays from an infinitely distant object to a point focus.

Let us suppose that the point S is in a medium of index of refraction 1 and the point S′ in a medium of index of refraction μ. Then the equation to the aplanatic surface becomes *

$$SP + \mu S'P = c,$$

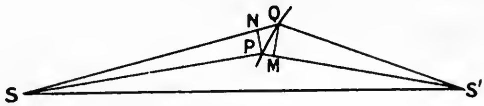


FIG. 20.

where P is any point on the surface. Let the point Q (fig. 20) be a neighbouring point to P. From P draw PN perpendicular to SQ and from Q draw QM perpendicular to S′P. Then, since the angle NSP is small we may write $SP = SN$; similarly $S'M = S'Q$. Now $SP + \mu S'P = c$ and $SQ + \mu S'Q = c$; hence by subtraction

$$SQ - SP = \mu(S'P - S'Q)$$

i.e. $$SQ - SN = \mu(S'P - S'M)$$

or $$NQ = \mu PM.$$

Divide both sides by PQ, which may be regarded as an element of a straight line. Then

$$\frac{NQ}{PQ} = \mu \frac{PM}{PQ}.$$

But $NQ/PQ = \cos NQP =$ sine (angle of incidence of SQ)
and $PM/PQ = \cos MPQ =$ sine (angle of refraction of PS′).
Therefore, in the limit when QP is made infinitely small we have sine (angle of incidence of SP) $= \mu$ sine (angle of refraction of PS′), i.e. the ray SP is refracted so as to pass through S′. Thus all rays incident on the surface from S pass after refraction through S′ and S′ is a real image of S.

Draw a circle with centre C and radius AC, A being any point on it; then with the same centre describe circles with radii equal to

* The curve represented by this equation is known as the Cartesian oval.

AC/μ and μAC, μ being greater than 1, and draw any line CP to meet them respectively in P and Q. Join QA and PA.

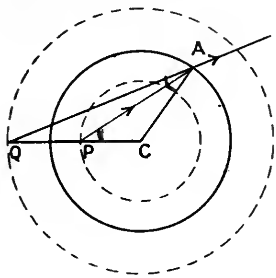


FIG. 21.

In $\triangle$s APC and AQC we have $\angle$ACP common, PC : CA : : CA : CQ, hence the $\triangle$s are similar and $\angle QAC = \angle APC$. Thus

$$\frac{\sin QAC}{\sin PAC} = \frac{\sin APC}{\sin PAC} = \frac{AC}{PC} = \mu.$$

If we regard the circle of radius CA as a section of a glass sphere with centre C and index of refraction μ, since AC is the normal to the surface at A, any ray diverging from the point P inside the sphere appears after refraction at the surface to come from Q. Thus Q is the image of P. P and Q are termed the aplanatic points of the sphere and the surface of the sphere is an aplanatic surface for P and Q. The sphere has an infinite number of aplanatic points; to every point on the inner sphere there corresponds an aplanatic point on the outer sphere.

To an element of area surrounding P on the inner sphere there corresponds as image an element of area surrounding Q on the outer sphere. The area of the whole inner sphere is $4\pi(AC/\mu)^2$ and the area of the whole outer sphere $4\pi(\mu AC)^2$. The area of the element surrounding Q is therefore μ^4 times the area of the element of which it is the image.

Ordinary lenses give sharp images only when the rays make small angles with the axis. The property of the aplanatic points of the sphere may be used to construct a lens which gives a sharp image, no matter at what angle the rays diverge from the axis.

For let ABED be the section of a lens of which CD is the axis; let P be the centre of curvature of the one surface DE and let P and Q be aplanatic points of the other surface AB. Then all rays diverging from P enter the lens undeviated and after refraction at the second surface appear to diverge from Q. The lens thus forms a virtual image Q of P.

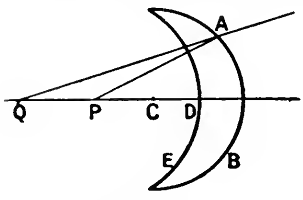


FIG. 22.

For the above construction to hold the object must of course occupy the one definite position and the light must be monochromatic.

Law of the Extreme Path. The laws of reflection and refraction can be summed up in a very general law entitled the law of the extreme path. It runs as follows: The optical length of the path traversed by a ray between two points is stationary, i.e. is either a maximum or a minimum. The law was formerly known as Fermat's Principle of Least Time, and was stated erroneously as follows: Rays of light are the shortest optical distance between the points they connect.

The law will first be proved for the case of a single reflection at a curved surface. This will of course include the case of reflection at a plane.

Let a ray of light from S fall on a curved surface AB at P and be reflected so as to arrive at S′. Take Q any other point on the curved surface, not necessarily in the same plane as S, P, and S′. Join SQ and S′Q. Then it is required to show that SP + S′P is either greater or less than SQ + S′Q for every position of Q whatever.

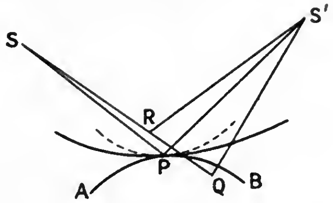


FIG. 23.

With S and S′ as foci draw an ellipsoid of revolution to touch the surface AB at P and let SQ intersect this ellipsoid at R. Then by the property of the ellipsoid of revolution $SP + S'P = SR + S'R$. But $S'R < RQ + S'Q$; hence $SR + S'R < SR + RQ + S'Q < SQ + S'Q$, i.e. $SP + S'P < SQ + S'Q$. If the surface AB had been more concave towards S and S′ than the ellipsoid itself is, e.g. like the dotted curve, then SQ would have intersected it before meeting the ellipsoid and we should consequently have $SP + S'P > SQ + S'Q$.

The case of refraction at a curved surface can be treated similarly by means of the other aplanatic surface.

Take next the case of two reflections. Let PABQ be the actual path of the ray and PA′B′Q a neighbouring path. We shall suppose that AA′ and BB′ are small quantities of the first order. Then the law of the extreme path demands that the difference between the two paths

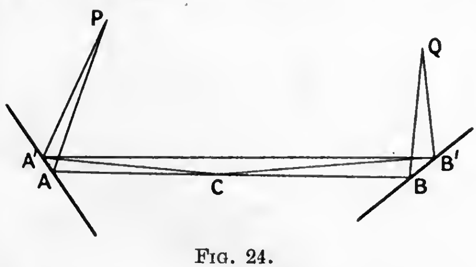


FIG. 24.

should be a small quantity of the second order. Take a point C on AB at some distance from both A and B. Then since the law of the extreme path holds for a single reflection, $PA' + A'C = PA + AC$ and $CB + BQ = CB' + B'Q$ as far as concerns small quantities of the first order. The path PA′CB′Q therefore differs from the path PACBQ only by small quantities of the second order. To complete the proof it is necessary to show that $A'C + CB'$ differs from $A'B'$ only by small quantities of the second order, and this is of course the case since the distance of C from A′B′ is first order.

More complicated cases can be treated in the same way.

EXAMPLES.

(1) If the light of the sun is admitted through a small hole, an image of the sun is formed on a screen placed to receive it, but if the aperture is a large one we obtain an image of the aperture. Explain this.

(2) The refractive indices of glass and water with respect to air are 1·52 and 1·33. Find the refractive index of glass with respect to water.

(3) A luminous point is placed between two mirrors inclined to one another at an angle of 90°. Find the number and position of the images formed by reflection at the mirrors, and draw a diagram showing the path to the eye of the axis of the pencil by which each image is seen.

(4) A luminous object is placed between two plane mirrors inclined to one another at an angle of 60°. Find the number of images produced and show that they all lie on a circle.

(5) Two plane mirrors inclined at an angle θ intersect in O; P is a point between the mirrors and PQR a ray emanating from P reflected at the mirrors in succession so as to return to P. Show that OP bisects $\angle$ QPR and that the length of the path is 2OP sin θ.

(6) On a moonlight night, when the surface of the sea is covered with small ripples, instead of a clear image of the moon a band of light is seen on the surface of the water extending in the direction of the moon. Explain with a diagram why this occurs.

(7) The apparent elevation of the centre of the sun's disc is 30°. Find the true elevation. The index of refraction for light passing from vacuum to air may be taken as 1·0003.

(8) To an observer looking down at a pool of clear water the depth appears to be 4 ft. What is the real depth?

(9) A small air bubble in a glass slab appears to an eye looking normally at the surface to be 2 cms. from the latter. If 1·52 is the index of refraction of the glass, what is the real distance of the bubble from the surface?

(10) A slab of glass 10 cms. thick with a refractive index of 1·52 is held with its lower surface 6 cms. above a piece of paper on which a mark is made. Where does the mark appear to be to an eye looking at it vertically through the slab? Illustrate your answer with a diagram.

(11) Two mirrors are inclined at a fixed angle to one another and the combination can be rotated about their line of intersection as axis. Show that, if a ray of light is reflected first in the one mirror and then in the other in a plane at right angles to the axis, the deviation is unaltered by the rotation of the mirror.

(12) Prove directly without using the property of aplanatic surfaces that, if a ray of light from a point P is reflected by a plane mirror so as to arrive at a point Q, then the optical length of the path of the ray is a minimum, i.e. is less than the optical length of any other path from P to the surface of the mirror and then to Q.

(13) Prove directly without using the property of aplanatic surfaces that, if a ray of light from a point P is refracted at the plane surface of a denser medium so as to arrive at a point Q inside that medium, then the optical length of the path of the ray is less than the optical length of any other path from P to Q.

(14) Deduce from the law of the extreme path, i.e. without using the law of reflection as an intermediate step, the formula connecting the positions of the object and image formed by reflection at a concave mirror.

CHAPTER II.

ELEMENTARY THEORY OF SPHERICAL MIRRORS AND LENSES.

Spherical Mirrors. If the surface bounding two media is spherical in shape and highly polished, it is said to form a spherical mirror. It is not necessary for it to be silvered; an unsilvered glass surface gives quite as sharp images, but if the glass surface is silvered or if the mirror is made of speculum metal, the images are much brighter. The glass must be silvered on the front if the simple theory given in this chapter is to apply, for if the mirror consists of a thin piece of glass silvered on the back, the light suffers refraction at the front, both before and after reflection at the back.

Spherical mirrors are divided into two classes, concave and convex. In the case of the concave spherical mirror, the light falls on the surface from the same side as the centre of curvature or centre of the sphere of which the surface forms part: in the case of the convex spherical mirror the light falls on the surface from the opposite side to the centre of curvature. Thus a glass-air surface can be regarded as concave or convex according to the side on which the source of light is placed.

The formulæ giving the positions of the images formed by mirrors and lenses are algebraic. Thus, if v denote the distance of an image formed by a mirror from the mirror, and we solve for v, we may find for an answer a value such as + 10 cms. or − 7 cms. It has hitherto been usual in elementary textbooks on light to take the direction opposed to the incident light as positive; thus $v = +10$ cms. means that the image is 10 cms. from the mirror on the side from which the light is coming, and $v = -7$ cms. means that the image is 7 cms. distant on the other side. In coordinate geometry and in plotting graphs on the other hand, positive numbers are always measured to the right and negative numbers to the left, and there is so much graph plotting done in schools now that this convention is very well understood. The two conventions agree or clash according to the side of the page from which the light comes.

In this book we shall adopt the convention of coordinate geometry throughout, and $v = -7$ cms. will denote that the image is 7 cms. to the left of the mirror, no matter from which side the light comes. There is no reason why a student should unlearn his coordinate geometry when he starts to study light, since the one convention possesses no advantage over the other.

Let AB represent a section of a concave spherical mirror, let C be its centre of curvature, and let P be a point source of light. Join PC and produce it to cut the mirror in A. We shall suppose at first that AP is greater than AC.

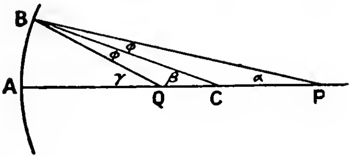


FIG. 25.

Draw any ray making a small angle α with AP to meet the mirror at B. Then CB is the normal to the element of the mirror at B. Let $\angle$PBC $= \phi$. The ray after reflection at B makes an angle ϕ with the normal and meets AP at Q. Denote $\angle$BCQ by β and $\angle$BQA by γ, and let the coordinates with respect to A of the points P, C, and Q have the values u, r, and v.

Then, from the figure

$$\beta = \phi + \alpha, \ \gamma = \phi + \beta$$

and consequently $\alpha + \gamma = 2\beta$. But since α, β and γ are small we can write

$$\alpha = \frac{\text{AB}}{u}, \ \beta = \frac{\text{AB}}{r}, \text{ and } \gamma = \frac{\text{AB}}{v}.$$

On substituting these values in the equation above and dividing out by AB we obtain

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r}.$$

If the point P is on the other side of C and the reflected ray meets AP to the right of A we proceed in the same way. If however BQ meets AP to the left of A (fig. 26) the equations between the angles are

$$\alpha = \phi + \beta, \ 2\phi = \alpha + \gamma,$$

which give $\alpha - \gamma = 2\beta$. In this case, however, $\gamma = -$ AB$/v$, so the

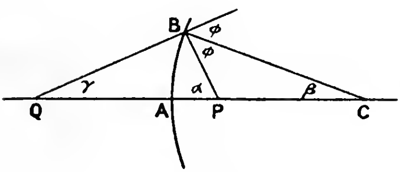


FIG. 26.

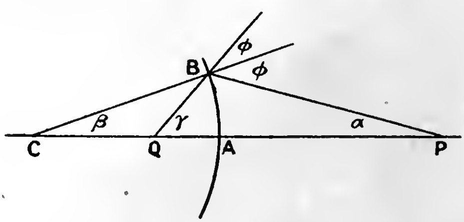


FIG. 27.

final equation is the same as before.

If the mirror is convex (fig. 27) the equations are

$$\gamma = \beta + \phi, \ 2\phi = \alpha + \gamma,$$

and these give $\gamma - \alpha = 2\beta$. In this case $\alpha =$ AB$/u$, $\beta = -$ AB$/r$, and $\gamma = -$ AB$/v$, so the equation again takes the same form. In figs. 26 and 27 it will be noticed that the reflected ray itself does not pass through Q but only its direction produced backwards.

Rays diverging from a point P therefore and making a small angle with the straight line joining P with the centre of curvature of the mirror, after reflection at the mirror either converge towards or appear to diverge from another point Q. Or in other words the mirror forms an image of the point P at Q.

In each of the three figs. 25, 26, and 27 keep the mirror and consequently the point C fixed and rotate the line PC through a small angle about an axis perpendicular to the plane of the figure through C. P and Q then describe arcs which may be taken as straight. Next rotate the figure about PC. Small circular elements of area will be traced out at P and Q and the one element will be the image of the other.

Spherical mirrors have usually a circular rim and the straight line through the centre of curvature perpendicular to the plane of this rim is called the axis of the mirror. We have proved, therefore, that a circular mirror always forms an image of a small plane figure situated on the axis and perpendicular to the axis.

If the object is at infinity, $1/u$ is zero and v becomes equal to $r/2$. All the incident rays are then parallel, and after reflection they converge to, or appear to diverge from, a point F midway between A and C. This point F is called the principal focus or simply the focus of the mirror and the length AF is called its focal length and is usually written f.

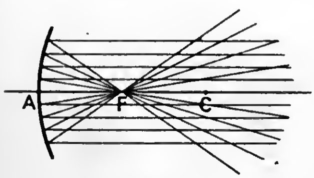


FIG. 28.

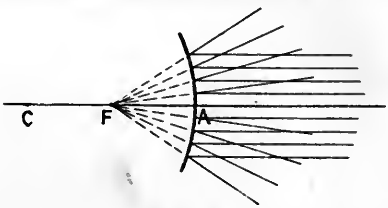


FIG. 29.

§ If the property of the focus and the fact that the mirror forms images be assumed, the position and size of the image can be found very easily by a graphical construction. Let there be an object at P (figs. 30, 31) only one-half of which, Pp, is drawn. Draw a ray from p through C; it falls on the mirror perpendicularly, and consequently after reflection comes back along the same path. Draw another ray pB parallel to the axis; after reflection it passes or appears to pass

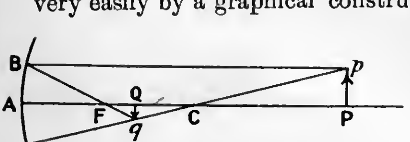


FIG. 30.

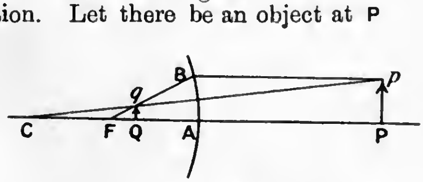


FIG. 31.

through the focus. The point q where BF cuts pC is therefore the image of p and the perpendicular to the axis, qQ, consequently the image of pP.

Since AB is small in comparison with the radius of curvature—it is greatly exaggerated in the figures for the sake of clearness—it may be regarded as straight and at right angles to AP. Then AB = Pp and consequently AB/Qq = Pp/Qq. Triangles ABF and QqF are similar as are also triangles QqC and PpC. Hence

$$\frac{\text{AB}}{\text{Q}q} = \frac{\text{AF}}{\text{QF}} \text{ and } \frac{\text{P}p}{\text{Q}q} = \frac{\text{PC}}{\text{QC}}.$$

But the left-hand sides of these two equations are already equal; hence

$$\frac{\text{AF}}{\text{QF}} = \frac{\text{PC}}{\text{QC}},$$

or

$$\frac{f}{f-v} = \frac{2f-u}{2f-v},$$

which gives

$$f(2f-v) = (2f-u)(f-v),$$

or

$$fu + fv = uv$$

and on dividing out by uvf

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} = \frac{2}{r},$$

the equation which has already been found.

The relative sizes of the image and object can also be obtained from figs. 30 and 31. For

$$\frac{\text{Q}q}{\text{P}p} = \frac{\text{QC}}{\text{PC}} = \frac{r-v}{-(u-r)} = \frac{v\left(\frac{1}{v}-\frac{1}{r}\right)}{-u\left(\frac{1}{r}-\frac{1}{u}\right)}.$$

But since

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r}, \qquad \frac{1}{v} - \frac{1}{r} = \frac{1}{r} - \frac{1}{u},$$

and

$$\frac{\text{Q}q}{\text{P}p} = -\frac{v}{u}.$$

The ratio Qq/Pp is termed the linear magnification. The minus sign means that if u and v have both the same sign then Qq and Pp are drawn in different directions from the axis, or, in other words, that the image is inverted. Apart altogether from the question of signs the above equation can be put into the following useful rule: the linear dimensions of the image and object are in the ratio of their distances from the mirror.

In working problems on mirrors it is better not to rely on the equation

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r} = \frac{1}{f}$$

alone, as errors of sign are very easily made, but to verify the result by the graphical construction.

As the position of the object varies, the change in the position and character of the image formed by a spherical mirror can be traced very readily by the graphical construction. It is given in the following tables:—

CONCAVE MIRROR.

Position of Object.	*Position of Image.*	*Character of Image.*
At infinity.	At focus.	Real.
Between ∞ and C.	Between F and C.	Real, inverted, diminished.
At C.	At C.	Real, inverted, same size.
Between C and F.	Between C and ∞.	Real, inverted, magnified.
At F.	At infinity.	
Between F and mirror.	From an infinite distance behind mirror to mirror.	Virtual, erect, magnified.
At mirror.	At mirror.	Erect, same size.

CONVEX MIRROR.

Position of Object.	*Position of Image.*	*Character of Image.*
At infinity.	At focus.	Virtual.
Between infinity and mirror.	Between F and mirror.	Virtual, erect, diminished.
At mirror.	At mirror.	Erect, same size.

§ An interesting optical illusion called the "Phantom Bouquet" can be produced with a large concave mirror of about a metre radius of curvature. A bouquet is placed in front of the mirror in a darkened room in such a position as to produce a real and magnified image. The bouquet itself is suspended upside down and the image is erect. The mirror and bouquet are placed in a large box and the bouquet illuminated strongly by wire filament lamps. These lamps are placed in the box in such a position that they cannot be seen from outside. After reflection by the mirror, the rays from the bouquet emerge from an opening in the box and form the image in front of it. An empty vase is placed below the image. If an observer looks at the vase so that his direction of vision meets the mirror, he sees the bouquet in the vase, but if he looks at it from the side, the vase appears empty.

A similar illusion, that has been shown at different places during the past four years, is one of dancing human figures about nine inches high. This also requires a darkened room for its exhibition. The figures are, of course, diminished images of real human beings, and their size is simply due to the image being much closer to the concave mirror than the object is. The axis of the concave mirror is at right angles to the line of vision of the spectators and the rays are turned towards the latter by a combination of plane mirrors, which at the same time inverts the image. The details of this combination are left as an exercise to the reader.

The illusion known as "Pepper's ghost" requires a large plane sheet of plate glass, which is placed vertical and at an angle of 45° with the line of vision of the spectators. The latter consequently see two scenes superimposed, the background which is visible by the direct rays that come through the glass, and objects at the side of the stage which are visible by the rays reflected through 90° by the glass. The background is usually somewhat dim, and at first the actor who is to fill the rôle of the ghost is kept in the dark at the side. He is too far to the side to be seen by the spectators directly. When it is time for him to "appear," the limelight is turned on him and his image at once appears superimposed on the background by the reflected light. The image is of course transparent; bright parts of the background can be seen through it. "Kineplastikon" is merely a variant of Pepper's ghost, in which the place of the live actor at the side is taken by a screen on which a picture of an actor is projected by the cinematograph.

Refraction at a Spherical Surface. Let us suppose that the common boundary of two transparent media is spherical in shape, and that the rays emitted by a luminous point P in the first medium are refracted at the spherical boundary. The first medium may be thought of as air and the second as glass, although for the sake of generality we shall put μ_1 for the index of refraction of the first medium and μ_2 for the index of refraction of the second, μ_2 being greater than μ_1.

Just as in the case of the spherical mirror the refracting surface may be either concave or convex with respect to the point P. We shall

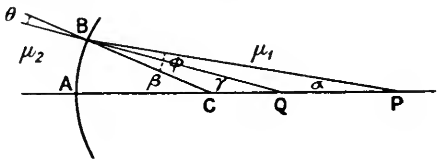


FIG. 32.

first take the case that it is concave (fig. 32).

Join P to C, the centre of curvature of the surface, and let PC meet the surface in A. Let PB be any ray making a small angle α with PC and meeting the surface at B. CB is the normal to the surface at B. After refraction the ray will proceed as if it came from Q. Let $\angle CBP = \phi$ and $\angle CBQ = \theta$. Then by the law of refraction $\mu_1 \sin \phi = \mu_2 \sin \theta$. Since the angle α is small, the angles ϕ and θ must also be small. Hence this equation may be written

$$\mu_1 \phi = \mu_2 \theta \qquad . \quad . \quad . \quad . \quad . \quad (2)$$

Let $\angle AQB = \gamma$ and $\angle ACB = \beta$. Then

$$\theta = \beta - \gamma, \quad \phi = \beta - \alpha,$$

and on substituting in equation (2) this gives

$$\mu_1(\beta - \alpha) = \mu_2(\beta - \gamma) \text{ or } \mu_2\gamma - \mu_1\alpha = \beta(\mu_2 - \mu_1) \qquad . \quad (3)$$

Now let $AP = u$, $AQ = v$, and $AC = r$, where u, v and r are all measured positive to the right of A. We may then write

$$\alpha = \frac{AB}{u}, \ \gamma = \frac{AB}{v}, \text{ and } \beta = \frac{AB}{r},$$

and on substituting these results in (3) and dividing out the common factor it reduces to

$$\frac{\mu_2}{v} - \frac{\mu_1}{u} = \frac{\mu_2 - \mu_1}{r} \qquad . \quad . \quad . \quad . \quad . \quad (4)$$

Imp.

If the point P is on the other side of C, $\phi = \alpha - \beta$ and $\theta = \gamma - \beta$, but the proof is otherwise the same. Thus all rays diverging from P and making a small angle with AP appear to come from Q after refraction, or in other words a virtual image of the point P is formed at Q.

If the refracting surface is convex (figs. 33 and 34) there are two cases according as BQ meets AP to the left or the right of A.

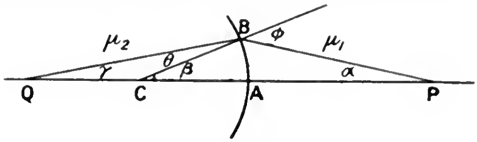


FIG. 33.

If we take the former case (fig. 33), $\phi = \alpha + \beta$, $\theta = \beta - \gamma$ and consequently

$$\mu_2\gamma + \mu_1\alpha = \beta(\mu_2 - \mu_1).$$

Since $\alpha = \mathrm{AB}/u$, $\gamma = -\mathrm{AB}/v$, and $\beta = -\mathrm{AB}/r$ this reduces to (4). In this case the image is real.

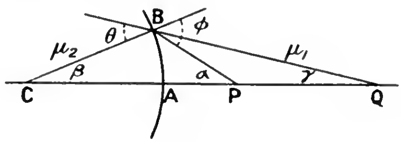


FIG. 34.

If we take the latter case (fig. 34), $\phi = \alpha + \beta$, $\theta = \beta + \gamma$, and therefore

$$\mu_2\gamma - \mu_1\alpha = \beta(\mu_1 - \mu_2).$$

In this case $\alpha = \mathrm{AB}/u$, $\gamma = \mathrm{AB}/v$, and $\beta = -\mathrm{AB}/r$, so the final equation is once more the same. In this case the image is virtual.

If in figs. 32, 33, and 34 the refracting surface and consequently C is kept fixed and PC rotated through a small angle about an axis through C perpendicular to the plane of the paper, P and Q describe small arcs which may be regarded as straight lines. If the figures then be rotated about AP these lines describe small areas, and every point on the one area is the image of a point on the other. Thus by refraction at a spherical surface, images are formed of plane elements on the axis of the surface and at right angles to the axis.

Refraction through a Lens. A lens is a portion of a refracting medium bounded by two spherical surfaces or by one spherical surface and a plane surface. The straight line joining the centres of curvature of the surfaces is called the axis of the lens, or, if one of the surfaces is plane the axis is the straight line normal to it drawn through the centre of curvature of the other. A plane through the axis is said to be a principal section of the lens.

Lenses are divided into two classes. The first class, convex or converging lenses, cause a beam of parallel rays to converge; the second class, concave or diverging lenses, cause a beam of parallel rays to diverge. Fig. 35 gives the principal sections of some typical lenses.

Of these, A, B, and C are convex; D, E, and F are concave. A is termed a double-convex or bi-convex lens, B a plano-convex lens, and C a convex meniscus, while D is termed a double-concave or bi-concave

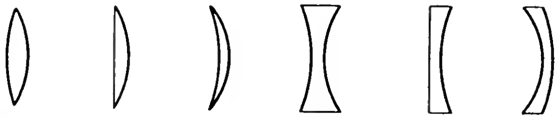


FIG. 35.

lens, E a plano-concave lens, and F a concave meniscus. C and F are termed also, somewhat indiscriminately, convexo-concave and concavo-convex lenses.

If a convex lens is placed in a medium the index of refraction of which is greater than that of the glass of the lens itself, it acts as a concave lens.

Let a lens be placed in a medium with reference to which the index of refraction of its material is μ and let a luminous point P be situated on its axis at a distance u from its first surface. Let r_1 and r_2 be the distances of the centres of curvature of the two surfaces each measured from its own surface, and let t be the thickness of the lens measured along its axis.

An image of P is formed by refraction at the first surface. Let its distance from the first surface be s. Then by (4)

$$\frac{\mu}{s} - \frac{1}{u} = \frac{\mu - 1}{r_1} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The image is distant $s + t$ from the second surface of the lens and a second image is formed of it by refraction at the latter. Let v be the distance of the second image from the second surface. Then applying (4) a second time

$$\frac{1}{v} - \frac{\mu}{s + t} = \frac{1 - \mu}{r_2} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Let us now suppose that the lens is thin and that t may be neglected in comparison with s. Equation (6) thus becomes

$$\frac{1}{v} - \frac{\mu}{s} = \frac{1 - \mu}{r_2} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Add the left-hand and right-hand sides of equations (5) and (7) and we obtain finally

$$\frac{1}{v} - \frac{1}{u} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad . \quad . \quad . \quad . \quad (8)$$

the fundamental equation giving the position of the image formed by a thin lens.

If the object is at an infinite distance from the lens, the incident rays are parallel, $1/u = 0$ and v is given by

$$\frac{1}{(\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right)}.$$

This quantity is denoted by f and is called the focal length of the lens. If f be substituted in (8) the fundamental equation of the thin lens becomes

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}.$$

If the lens is convex, the image formed of the object at infinity is

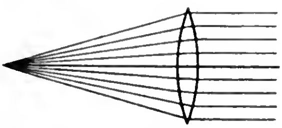


FIG. 36.

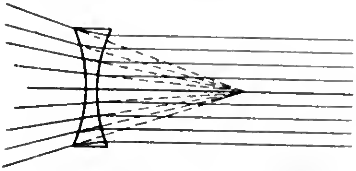


FIG. 37.

situated on the other side of the lens from the incident light and is real; if the lens is concave, it is situated on the same side of the lens as the incident light and is virtual.

In using the equation for f,

$$\frac{1}{f} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right),$$

due regard must be paid to the signs of r_1 and r_2. If the lens is bi-convex or bi-concave the surfaces are turned different ways and r_1 and r_2 have different signs. Which is positive and which is negative depends upon whether the lens is bi-convex or bi-concave and also on which side of the lens is to the right of the page, but in any case, after the numerical values are substituted, the two terms in the second bracket must have the same sign. If the lens is concavo-convex or convexo-concave r_1 and r_2 have the same sign, and after the numerical values are substituted the two terms in the second bracket have consequently different signs; the two curved surfaces partly neutralize one another. If the lens is plano-convex or plano-concave, either r_1 or r_2 becomes infinite.

Optical Centre of a Lens. If a ray of light passes through a lens undeviated, that is, if its direction after emergence is parallel to its direction before incidence, the lens must act on it merely as a parallel slab and the tangent planes at the points of incidence and emergence must be parallel.

Let $E_1G_1G_2E_2$ be such an undeviated ray. Then the radii C_1G_1 and C_2G_2 perpendicular to the tangent planes must be parallel. Let G_1G_2 cut the axis of the lens at C.

Triangles C_1CG_1 and C_2CG_2 are similar, for $\angle G_1C_1C = \angle G_2C_2C$, since

C_1G_1 is parallel to C_2G_2, and $\angle C_1CG_1 = \angle C_2CG_2$. Thus $C_1C : C_2C :: C_1G_1 : C_2G_2$, or, in other words, the distance between the centres is

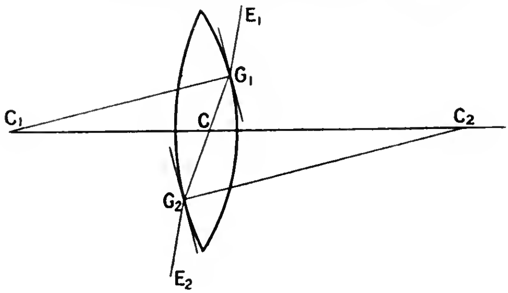


FIG. 38.

divided in the ratio of the radii and C is a fixed point. All rays, therefore, that are undeviated by the lens pass through a fixed point on the axis. This point is called the optical centre of the lens.

Conversely, if any ray passes through the optical centre, it is undeviated, for, if the radii be drawn to the points of incidence and emergence, they can be proved to be parallel.

Fig. 38 is exaggerated for the sake of clearness. It will never be necessary to consider rays incident so obliquely as E_1G_1. When the lens is thin, we may regard its optical centre and the two points in which the axis meets its surface as coincident.

Graphical Determination of Position of Image. If the properties of the focus and the optical centre of the lens be assumed, the position of the image can be determined very easily graphically. For, let Pp represent one-half of the object. From p draw a ray parallel to the axis to meet the lens in B. After refraction by the lens it either passes through F (fig. 39) or appears to pass through F (fig. 40). From p draw another ray through C, the centre of the lens.

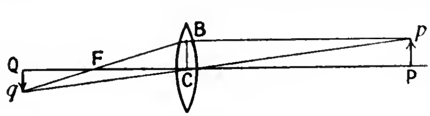


FIG. 39.

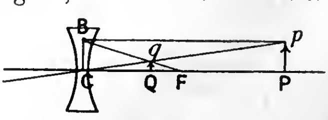


FIG. 40.

It passes through undeviated. The point where these two rays intersect, that is, q, is the image of p and the perpendicular to the axis, qQ, is the image of pP.

The fundamental formula for the thin lens can be derived very easily from figs. 39 and 40. For CB = Pp and consequently CB/Qq = Pp/Qq. Triangles FQq and FCB are similar as are also triangles CQq and CPp. Hence

$$\frac{CB}{Qq} = \frac{CF}{QF} \text{ and } \frac{Pp}{Qq} = \frac{PC}{QC}.$$

But the left-hand sides of these two equations are already equal; hence

$$\frac{CF}{QF} = \frac{PC}{QC} \text{ or } \frac{f}{f-v} = \frac{-u}{-v},$$

which gives $vf = uf - uv$, and on dividing out by uvf

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}.$$

The relative sizes of the image and object are also obtained from figs. 39 and 40.

For $$\frac{Qq}{Pp} = \frac{CQ}{CP} = \frac{v}{u},$$

that is, the linear dimensions of the image and object are in the ratio of their distances from the lens.

In working problems on lenses it is better not to rely on the equation

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}$$

alone, as errors of sign are very easily made, but the result should be verified by the graphical construction.

As the position of the object varies, the change in the position and character of the image formed by a lens can be traced very readily by the graphical construction. It is given in the following tables:—

CONVEX LENS.

Position of Object.	*Position of Image.*	*Character of Image.*
At infinity.	At focus.	Real.
Between ∞ and $u = -2f$.	Between $v = f$ and $v = 2f$.	Real, inverted, diminished.
At $u = -2f$.	At $v = 2f$.	Real, inverted, same size.
Between $u = -2f$ and $u = -f$.	Between $v = 2f$ and $v = -\infty$.	Real, inverted, magnified.
At $u = -f$.	At $v = -\infty$.	
Between $u = -f$ and lens.	Between $v = +\infty$ and lens.	Virtual, erect, magnified.

CONCAVE LENS.

Position of Object.	*Position of Image.*	*Character of Image.*
At infinity.	At focus.	Virtual.
Between ∞ and lens.	Between focus and lens.	Virtual, erect, diminished.

WORKED EXAMPLES.*

(1) A concave mirror has a radius of curvature of 16 cms. Find the position, nature, and size of the image when an object 5 mm. high is placed (*a*) 20 cms. from the mirror, (*b*) 6 cms. from the mirror.

(*a*) The formula gives

$$\frac{1}{20} + \frac{1}{v} = \frac{2}{16}.$$

Hence $v = 13\frac{1}{3}$ cms. and the image is real and inverted. The magnification v/u is $\frac{2}{3}$ and hence the height of the image $3\frac{1}{3}$ mm.

* In each of the following examples it has been assumed that the light is coming from the right.

(*b*) In this case

$$\frac{1}{6} + \frac{1}{v} = \frac{2}{16}.$$

Hence $v = -24$, i.e. the image is behind the mirror and is erect and virtual. The magnification is $-\frac{24}{6} = -4$ and the height of the image consequently 2 cms.

(2) A convex mirror has a radius of curvature of 16 cms. Find the position, nature, and size of the image when an object 5 mm. high is placed at a distance of 20 cms. from the mirror.

In this case the centre of curvature is behind the mirror so that r must be written $= -16$. Then the formula gives

$$\frac{1}{20} + \frac{1}{v} = -\frac{2}{16}.$$

Hence $v = -5\frac{5}{7}$, the image is behind the mirror, erect and virtual, the magnification is $-\frac{40}{7 \times 20} = -\frac{2}{7}$ and the height of the image is $\frac{10}{7}$ mm.

(3) The radii of curvature of the two faces of a thin convex lens are 20 and 40 cms. and the refractive index of the glass of which it is made is 1·5. Find the numerical value of the focal length, (*a*) if the lens is biconvex, (*b*) if it is concavo-convex.

We have in the first case

$$\frac{1}{f} = (1·5 - 1)\left(\frac{1}{20} + \frac{1}{40}\right), \text{ i.e. } f = 26\tfrac{2}{3} \text{ cms.}$$

and in the second case

$$\frac{1}{f} = (1·5 - 1)\left(\frac{1}{20} - \frac{1}{40}\right) \text{ or } f = 80 \text{ cms.}$$

(4) The focal length of a convex lens is 40 cms. Find the positions of the image when the object distance is (*a*) 60 cms., (*b*) 30 cms.

Here f is written -40 since rays are brought to a focus on the negative side of the lens. In the first case the formula is

$$\frac{1}{v} - \frac{1}{60} = -\frac{1}{40},$$

$v = -120$ cms., and the image is a real one and is on the opposite side of the lens from the object. In the second case

$$\frac{1}{v} - \frac{1}{30} = -\frac{1}{40},$$

$v = +120$ cms., the image is a virtual one and is on the same side of the lens as the object is.

EXAMPLES.

(1) The radius of curvature of a concave mirror is 20 cms. Find the positions of the object for which a real image three times its height and a virtual image twice its height are formed.

(2) A flower is suspended, inverted, 50 cms. in front of a concave mirror. An empty vase is placed 100 cms. in front of the mirror. What must be the radius of curvature of the mirror, if the flower appears in the vase to an observer looking at the vase towards the mirror?

(3) A glass globe, 6 inches in diameter, is filled with water. Trace the changes in position of the image, seen by an observer looking along a

diameter, of a point in the water as it moves from the farther to the nearer end of the diameter. The thickness of the glass may be neglected.

(4) The radius of curvature of a convex mirror is 15 cms. The object is distant 1 metre and has a height of 5 cms. Find the position and height of the image.

(5) A small air bubble in a sphere of glass of 3 cms. diameter appears to be 1 cm. from the surface when looked at along a diameter. If the refractive index of the glass is 1·52, find the true position of the bubble.

(6) An incandescent gaslight with a mantle 8 cms. high stands at the same level as a convex lens of focal length 20 cms. and at a distance of 5 metres from it. Find the position of the image. If the light is lifted 50 cms. above its original position, what change takes place in the position of the image?

(7) Determine by experiment the form of a curve showing the relation between the distance of an object from a given lens and the magnification of the real image.

(8) An image of a lamp is to be produced on a screen at a distance of 4 metres and is to be magnified four times. Find the focal length of the lens that will be required.

(9) A lens of focal length 12 cms. made of glass of refractive index 1·52 is immersed in water. What does the focal length become?

(10) A convex lens of focal length f produces a real image of magnification m. Show that the distance of the object from the lens must be $(m + 1)f/m$.

(11) An image of height a is formed on a screen by a convex lens. It is found by keeping the positions of object and screen unaltered and by moving the lens towards the screen, that there is a second position for the lens in which it forms a sharp image on the screen. In this case the height of the image is b. Show that the height of the object is $\sqrt{ab}$.

(12) Prove that the distance between two real conjugate points with respect to a convex lens cannot be less than four times its focal length.

(13) The radii of curvature of a biconvex lens are 20 and 30 cms. and it is made of glass of refractive index 1·52. Calculate its focal length. If the lens had been a convex meniscus lens with the same radii of curvature, what would the focal length have been?

(14) If a plane mirror on which a pencil of light is incident is turned through any angle about an axis perpendicular to the plane of incidence, the reflected light is deviated through twice that angle.

Hence show that when a plane wave is reflected at a spherical surface, the curvature of the reflected wave is twice that of the surface.

(15) Prove that if a lens is held before the eye and then moved to one side, the objects seen through it appear to move in the same direction as the lens if the latter is concave, but in the opposite direction to the lens if the latter is convex.

(16) OA and OB are two straight lines at right angles to one another. A point C is taken in OA and a point D is taken in OB, so that OC = OD = f, and straight lines are drawn through D and C parallel to OA and OB to meet in E. A straight line is drawn through E to meet OA and OB in P and Q. Show that if OP represents the object distance for a convex lens of focal length f, OQ represents the image distance.

Show that the construction holds for a concave lens, if C and D are taken in OA and OB produced.

(17) The focal length of a thin lens has been determined to be 25 cms. by using a lamp at a great distance as object and measuring the distance of

the image from the lens. How far away must the lamp be in order that the result may be right to 3 per cent.?

(18) A system of rays is such that they all cut a given surface orthogonally. A curved mirror intersects each ray in a point, such that the sum or difference of the distances of the point of intersection from the orthogonal surface and from a fixed point is constant. Show that the mirror reflects the system of rays to a focus at the fixed point.

(19) A convex and a concave lens each 20 cms. in focal length are placed coaxially at a distance of 6 cms. apart. Find the position of the image, if the object is at a distance of 30 cms. beyond (*a*) the convex, (*b*) the concave lens.

CHAPTER III.

THICK LENSES AND SYSTEMS OF LENSES.

If we are dealing with a thin lens a knowledge of its focal length enables us to calculate the distance of the image when the distance of the object is known. It is immaterial from what point in the lens these distances are measured. If, however, the lens is a thick one or a photographic objective consisting of four or six separate lenses, very different values of the object and image distances are obtained according as they are measured from the front or the back surface of the lens, and the question arises, are there any points from which they can be measured in this case in order to give consistent results, or must the single surfaces be treated separately.

It was proved by Gauss seventy years ago that it is not necessary to treat the single surfaces separately, but that a compound lens can be treated as a whole and the ordinary formula for the thin lens applies, if the object and image distances are measured from two theoretical planes fixed with reference to the lens. When the medium on both sides of the lens is the same, these planes are called the equivalent planes; when it is different they are called the principal planes. The rays from the object diverge to the one equivalent plane and converge from the other to the image, conjugate rays meeting the equivalent planes the same distance from the axis.

It is always possible to find a thin lens that produces an image of a given object in the same place and of the same size as the image produced by a system of lenses. For this only means the solution of the equations

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}, \quad m = v/u, \quad d = u - v,$$

where d and m are given. But the equivalence holds only for the two given conjugate points. The images of other objects produced in the two ways coincide neither in position nor magnitude. It is thus impossible to find a single lens that placed in any one position will act in the same way as the system of lenses does.

If, however, a thin lens of a certain focal length is taken and placed at the first equivalent plane to receive the rays of light and then rapidly shifted to the other equivalent plane to discharge the rays, it will act in exactly the same manner as the compound lens.

This is, of course, an exceedingly popular way of putting the matter, but it will serve to give preliminary ideas.

In this chapter the whole treatment is made to depend on the property of conjugate planes and on Helmholtz's law of magnification. This method does not enable the positions of the cardinal points to be calculated in the general case, but it is much simpler than the other methods and it is at the same time rigorous. Possibly the analytical nature of the other methods has been the reason why the theory is so little known at present.

Helmholtz's Law. Let Qq be the image of Pp formed by a refracting surface (figs. 41, 42, 43, cf. figs. 32, 33, 34). Let PB be any ray through P making a small angle with the axis and let BQ be the

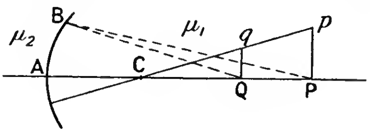


FIG. 41.

same ray or its direction produced backwards after refraction. Let y_1 = Pp and y_2 = Qq, let a_1 and a_2 be the angles which PB and QB

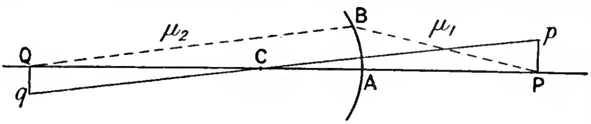


FIG. 42.

make respectively with the positive direction of the axis of the lens. We shall regard y_1 and y_2 as the coordinates of p and q; they are

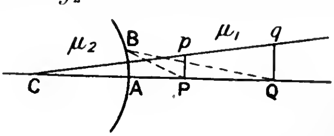


FIG. 43.

positive if the points are above the axis and negative if the points are below the axis. Then

$$\frac{y_1 \tan a_1}{y_2 \tan a_2} = \frac{\mathrm{P}p\,\dfrac{\mathrm{AB}}{\mathrm{AP}}}{\mathrm{Q}q\,\dfrac{\mathrm{AB}}{\mathrm{AQ}}} = \frac{\mathrm{PC}}{\mathrm{QC}}\frac{\mathrm{AQ}}{\mathrm{AP}} = \frac{(r-u)v}{(r-v)u} \quad . \quad . \qquad (9)$$

But
$$\frac{\mu_2}{v} - \frac{\mu_1}{u} = \frac{\mu_2 - \mu_1}{r},$$

that is, $\mu_2\left(\frac{1}{v}-\frac{1}{r}\right)=\mu_1\left(\frac{1}{u}-\frac{1}{r}\right)$ or $\mu_2 u\,(r-v)=\mu_1 v\,(r-u)$.

Thus, on substitution in (9), we obtain

$$\frac{y_1 \tan \alpha_1}{y_2 \tan \alpha_2}=\frac{\mu_2}{\mu_1}$$

or

$$\mu_1 y_1 \tan \alpha_1 = \mu_2 y_2 \tan \alpha_2 \quad . \quad . \quad . \quad . \quad (10)$$

This formula was first given by Lagrange. The ratio y_2/y_1 is called the linear magnification and the ratio $\tan \alpha_2/\tan \alpha_1$ the angular magnification for the conjugate points P and Q.

Similarly it may be shown that in the case of reflection at a spherical mirror

$$y_1 \tan \alpha_1 = -\,y_2 \tan \alpha_2.$$

Suppose now that instead of one spherical refracting surface separating two media we have $n-1$ coaxial refracting surfaces separating n media, and that the index of refraction of the first medium is μ_1, of the second medium μ_2, and of the n^{th} medium μ_n. Let a small object of linear dimensions y_1 perpendicular to the axis be placed on the axis in the first medium and let a ray drawn from it make an angle α_1 with the axis. After refraction at the first surface this ray makes an angle α_2 with the axis and appears to diverge from an image of linear dimensions y_2. After refraction at the second surface it makes an angle α_3 with the axis and appears to diverge from an image of linear dimensions y_3. Thus, applying (10), we obtain

$$\mu_1 y_1 \tan \alpha_1 = \mu_2 y_2 \tan \alpha_2,$$
$$\mu_2 y_2 \tan \alpha_2 = \mu_3 y_3 \tan \alpha_3,$$
$$. \quad . \quad . \quad . \quad . \quad . \quad .$$
$$\mu_{n-1}\, y_{n-1} \tan \alpha_{n-1} = \mu_n y_n \tan \alpha_n,$$

or leaving out the intermediate steps

$$\mu_1 y_1 \tan \alpha_1 = \mu_n y_n \tan \alpha_n \quad . \quad . \quad . \quad (11)$$

This equation for a system of surfaces was first given by Helmholtz.

Focal Planes. Let us suppose we have any system of coaxal spherical refracting surfaces, that a plane element is placed on the axis at a point P and that a final image is formed of this element by the system at a point Q. Take any point on the axis as origin and let x_1 and x_2 be the coordinates of P and Q with respect to this origin. We shall have no occasion to deal with the intermediate images and so can give the suffix 2 to the last image. Then

$$\mathrm{A}x_1x_2 + \mathrm{B}x_1 + \mathrm{C}x_2 + \mathrm{D} = 0 \quad . \quad . \quad . \quad (12)$$

where A, B, C, and D are constants depending on the positions and curvature of the surfaces and the indices of refraction of the different media. This equation merely expresses the fact that for every position of the object there is always one and only one position of the final image, a fact which we can prove by taking the single refractions separately, and it is the most general way of expressing this fact, as may be seen on attempting to add additional terms.

For a single refraction we have

$$\frac{\mu_2}{v} - \frac{\mu_1}{u} = \frac{\mu_2 - \mu_1}{r}.$$

On differentiating with respect to u this gives

$$-\frac{\mu_2}{v^2}\frac{dv}{du} + \frac{\mu_1}{u^2} = 0$$

or

$$\frac{dv}{du} = \frac{\mu_1 v^2}{\mu_2 u^2}$$

and this is always positive. Thus, if the object is displaced along the axis, the image is displaced along the axis in the same direction. Since this holds true of all the intermediate refractions it must be true for the final image formed by a system.

Differentiate (12) with regard to x_1. Then

$$Ax_2 + Ax_1\frac{dx_2}{dx_1} + B + C\frac{dx_2}{dx_1} = 0$$

and

$$\frac{dx_2}{dx_1} = -\frac{Ax_2 + B}{Ax_1 + C}$$

$$= \frac{A\dfrac{Bx_1 + D}{Ax_1 + C} - B}{Ax_1 + C} = -\frac{(BC - AD)}{(Ax_1 + C)^2}.$$

Since this must be positive and the denominator being a square is always positive, it follows that BC − AD is negative.

Assume that A is not zero. Then (12) may be written in the form

$$x_1x_2 + \frac{B}{A}x_1 + \frac{C}{A}x_2 + \frac{BC}{A^2} = \frac{BC}{A^2} - \frac{D}{A}$$

or

$$(x_1 - g_1)(x_2 - g_2) = -\gamma^2$$

where $g_1 = -C/A$, $g_2 = -B/A$, and $-\gamma^2$ is written for $(BC - AD)/A^2$ since the latter is essentially negative.

The region of space, in which all possible positions of the object are situated, is called the object space and the region in which all possible positions of the image are situated is called the image space. The object and image spaces may, of course, overlap.

If the object P is at infinity, x_1 is infinite and $x_2 - g_2$ must equal zero, that is, the image is situated on the plane $x_2 = g_2$. This plane is called the focal plane of the image space. Similarly, if the image is at infinity, the object is situated on the plane $x_1 = g_1$ and the latter is called the focal plane of the object space.

Principal Planes. Nodal Points. Focal Lengths of System. Let U and V denote the distances of the object and image from the focal planes. Then

$$U = x_1 - g_1,\ V = x_2 - g_2.$$

Let Pp be a linear object of length y_1, Qq its image of length y_2, PQ the axis of the system, BF_1 and AF_2 the lines in which the focal planes of the object space and image space respectively intersect the plane of the diagram.

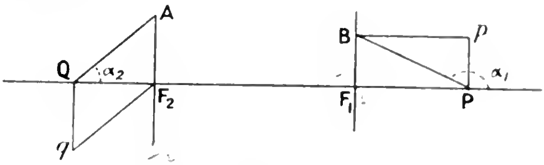


FIG. 44.

Draw pB parallel to PF_1. Then since pB is parallel to the axis in the object space, after passing through the system it must intersect the focal plane of the image space in F_2. It must also pass through q since the point q is the image of the point p. Consequently F_2q is the image of the ray pB after refraction through the system, or, in other words, the ray F_2q is conjugate to the ray pB.

Draw QA parallel to qF_2. The ray conjugate to QA in the object space must pass through P, the conjugate point to Q. It must also intersect BF_1 in the same point as the ray conjugate to qF_2, that is, as Bp does. It is therefore BP.

We can now apply Helmholtz's law. $F_1P = U$, $F_2Q = V$. $\tan \alpha_1 = -F_1B/F_1P = -y_1/U$, $\tan \alpha_2 = F_2A/QF_2 = y_2/V$. Let μ_1 be the index of refraction of the medium in which the object is placed and μ_2 the index of refraction of the medium in which the image is placed. Then

$$-\frac{\mu_1 y_1^2}{U} = \frac{\mu_2 y_2^2}{V}$$

or

$$\frac{y_2^2}{y_1^2} = -\frac{\mu_1 V}{\mu_2 U} = \frac{\mu_1 \gamma^2}{\mu_2 U^2} = \frac{\mu_1 V^2}{\mu_2 \gamma^2} \quad . \quad . \quad . \quad (13)$$

since $UV = -\gamma^2$.

If the linear magnification is unity, $y_2 = y_1$, and by substitution in (13)

$$\left.\begin{aligned} U^2 &= \frac{\mu_1}{\mu_2}\gamma^2 = f_1^2 \\ V^2 &= \frac{\mu_2}{\mu_1}\gamma^2 = f_2^2 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (14)$$

the above equations serving as the definitions of f_1^2 and f_2^2. On taking the root of equations (14) we obtain $U = \pm f_1$, $V = \pm f_2$, two object positions and two image positions. If we had substituted $y_2 = -y_1$ in equation (13) we would have had the same result. One pair of the conjugate points found obviously gives the magnification unity and the other the magnification minus unity. Which is which depends on the absolute signs of f_1 and f_2, quantities the squares of which have alone hitherto been defined.

The pair of points which give the magnification unity is called the principal points. The other pair is called the anti-principal points. The planes through the principal points and at right angles to the axis are called the principal planes. They are such that if any ray cuts the one a certain distance from the axis, its conjugate cuts the other the same distance from the axis.

Let us now define the absolute values of f_1 and f_2 by stating that the principal planes of the object and image spaces are given respectively by

$$\mathrm{U} = -f_1, \ \mathrm{V} = -f_2.$$

Then, since

$$\mathrm{UV} = -\gamma^2, \ f_1 f_2 = -\gamma^2,$$

and by substituting this value for γ^2 in (14) we find $\mu_2 f_1 = -\mu_1 f_2$.

Let $\alpha_1 = \alpha_2$. Then $\tan \alpha_1 = \tan \alpha_2$ and by Helmholtz's law $\mu_1 y_1 = \mu_2 y_2$. We can rewrite (13)

$$\frac{y_2^2}{y_1^2} = \frac{f_1^2}{\mathrm{U}^2} = \frac{\mathrm{V}^2}{f_2^2},$$

or, taking the root,

$$\frac{y_2}{y_1} = -\frac{f_1}{\mathrm{U}} = -\frac{\mathrm{V}}{f_2} \quad . \quad . \quad . \quad . \quad . \quad (15)$$

the sign being now no longer ambiguous. Multiplying the numerators by μ_2 and the denominators by μ_1 we obtain

$$\frac{\mu_2 y_2}{\mu_1 y_1} = -\frac{\mu_2 f_1}{\mu_1 \mathrm{U}} = -\frac{\mu_2 \mathrm{V}}{\mu_1 f_2},$$

or

$$\frac{\mu_2 y_2}{\mu_1 y_1} = \frac{f_2}{\mathrm{U}} = \frac{\mathrm{V}}{f_1},$$

since $\mu_2 f_1 = -\mu_1 f_2$. Hence, if $\alpha_1 = \alpha_2$,

$$\mathrm{U} = f_2, \ \mathrm{V} = f_1.$$

The points defined by the above equation are called the nodal points. They are two points on the axis, such that if the incident ray passes through the one the emergent ray passes through the other and further is parallel to the incident ray.

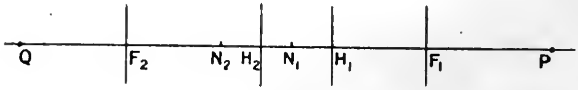


FIG. 45.

In fig. 45 let Q be the image of P, let F_1, H_1 be the focal and principal planes of the object space and F_2, H_2 the focal and principal planes of the image space. Then $\mathrm{F_1P} = \mathrm{U}$ and $\mathrm{F_2Q} = \mathrm{V}$. Let N_1 be the nodal point of the object space, and N_2 the nodal point of the image space. By definition $\mathrm{H_1F_1} = f_1$, $\mathrm{H_2F_2} = f_2$, $\mathrm{F_1N_1} = f_2$, and $\mathrm{F_2N_2} = f_1$.

Let u and v be the coordinates of P and Q with reference to their respective principal planes. Then

$$u = H_1P = f_1 + U$$

and

$$v = H_2Q = f_2 + V.$$

On substituting from the above for U and V in $UV = -\gamma^2$ and remembering that $-\gamma^2 = f_1 f_2$ we obtain

$$(u - f_1)(v - f_2) = f_1 f_2,$$

which simplifies to

$$\frac{f_2}{v} + \frac{f_1}{u} = 1 \quad . \quad . \quad . \quad . \quad . \quad (16)$$

f_1 and f_2 are called the focal lengths of the object and image space respectively.

If the initial and final media are the same, $\mu_1 = \mu_2$ and $f_2 = -f_1$. Write in this case $f_2 = -f_1 = f$. Then equation (16) becomes

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f},$$

which is exactly the same as the fundamental equation for the thin lens.

When $f_2 = -f_1 = f$ the nodal points are in the principal planes, the principal planes are called the equivalent planes of the system and f is called the equivalent focal length of the system.

Expression for the Magnification. According to (15)

$$\frac{y_2}{y_1} = -\frac{V}{f_2}.$$

On substituting $v - f_2$ for V this gives

$$\frac{y_2}{y_1} = -\frac{(v - f_2)}{f_2} \quad . \quad . \quad . \quad . \quad . \quad (17)$$

But from (16) $\dfrac{v - f_2}{f_2} = \dfrac{vf_1}{uf_2}$.

Substitute in (17); then $\dfrac{y_2}{y_1} = -\dfrac{vf_1}{uf_2}$.

Since $\mu_2 f_1 = -\mu_1 f_2$ this gives

$$\frac{y_2}{y_1} = \frac{\mu_1 v}{\mu_2 u}.$$

If the initial and final media are the same, this reduces to $y_2/y_1 = v/u$, i.e. the magnification is equal to the ratio of the distances of the image and object from their respective principal planes.

Graphical Construction of Images. The theory of the foregoing pages was first given by Gauss, but not in the same way as is done here. The nodal points were introduced by Listing. The foci, principal points, and nodal points are referred to as the Gauss points or cardinal points of a lens or system of lenses. Strictly speaking, the properties of these points hold true only when images are formed of small objects on the axis by rays inclined at a small angle to the axis, but they are important as giving the first approximation to the theory of the majority of optical instruments in use.

No matter of how many lenses a system of coaxial spherical refracting surfaces may consist or how many media of different refractive indices the rays may travel through, if the position of the object is given and the cardinal points of the system are known, the position and magnification of the image can at once be deduced. It is not necessary to know anything at all about the position and curvature of the refracting surfaces or the nature of the intermediate media the rays travel through. A knowledge of the cardinal points is sufficient.

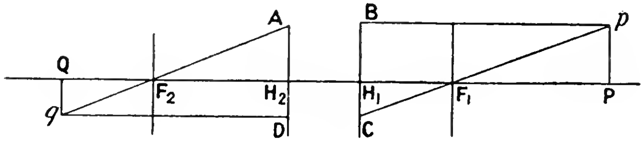


FIG. 46.

For example in fig. 46 let Pp be an object, F_1, F_2, H_1 and H_2 the focal planes and principal planes, and let it be required to find the position and size of the image.

Through p draw a ray parallel to the axis to meet the principal plane in B. Then, since pB is parallel to the axis its conjugate must pass through F_2, and it must meet the principal plane of the image space in A, A being the same distance from the axis as B. It is thus fully determined. Join pF_1 and produce it to meet the principal plane in C. Take a point D on the other principal plane the same distance from the axis that C is, and draw a ray qD through it parallel to the axis. This ray is the conjugate of Cp, for since Cp goes through the focus in the object space, its conjugate in the image space must be parallel to the axis. Thus q is fully determined.

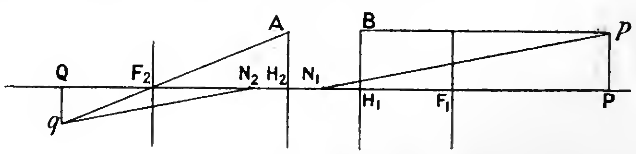


FIG. 47.

Fig. 47 shows another way of proceeding. The ray Aq conjugate to pB is found in the same way as before. Then a straight line is drawn joining p to N_1, the nodal point of the object space, and through N_2, the nodal point of the image space, N_2q is drawn parallel to pN_1. The point in which it cuts AF_2, that is, q, is the image of p.

The cardinal points may be obtained either by experiment or calculation. They do not require to occur in any particular order.

In what follows we shall confine our attention chiefly to systems, in which air is the initial and final medium, and consequently shall have usually to speak of the equivalent planes instead of the principal planes and nodal points of the system.

The Thick Lens. Let $E_1G_1G_2E_2$ be a ray which passes through a thick lens undeviated. Let C be the optical centre of the lens (cf. p. 27) and C_1, C_2 the centres of curvature of its two surfaces. Produce E_1G_1 to meet the axis in N_1 and produce E_2G_2 to meet the axis in N_2.

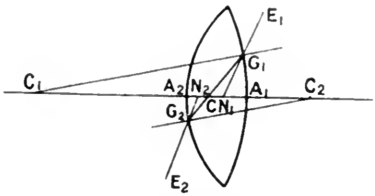


FIG. 48.

In $\triangle C_1G_1N_1$ $\sin C_1G_1N_1 : \sin G_1C_1C :: C_1N_1 : G_1N_1$. In $\triangle C_1G_1C$ $\sin C_1G_1C : \sin G_1C_1C :: C_1C : CG_1$. Hence, eliminating $\sin G_1C_1C$ we obtain

$$\frac{\sin C_1G_1N_1}{\sin C_1G_1C} = \frac{G_1C}{G_1N_1}\frac{C_1N_1}{C_1C}.$$

But $\angle C_1G_1N_1$ is equal to the angle of incidence at G_1 and $\angle C_1G_1C$ is equal to the angle of refraction at G_1. Hence

$$\frac{\sin C_1G_1N_1}{\sin C_1G_1C} = \mu,$$

where μ is the index of refraction of the glass of the lens.

Let the point G_1 move down the surface of the lens to A_1. As it does so, the point N_1 moves along the axis. The limiting position of N_1 when G_1 reaches A_1 is the nodal point of the object space. We have

$$\mu = \frac{G_1C}{G_1N_1}\frac{C_1N_1}{C_1C}.$$

In the limit this becomes

$$\mu = \frac{A_1C}{A_1N_1}\frac{C_1N_1}{C_1C}.$$

If the radius of curvature is great in comparison with the thickness of the lens C_1N_1 may be put equal to C_1C. Then

$$\mu = \frac{A_1C}{A_1N_1}.$$

Similarly it may be shown that

$$\mu = \frac{A_2C}{A_2N_2}.$$

Thus in a symmetrical double-convex lens of index of refraction 1·52 the nodal points and consequently the equivalent planes are situated inside the lens approximately one-third of its thickness from each surface. The same holds true for a symmetrical double-concave lens.

In the case of a plano-convex or plano-concave lens of index of refraction μ one of the nodal points is obviously in the point where the axis meets the curved surface, and coincides with the optical centre of the lens. If t is the thickness of the lens, the other nodal point is in the lens and situated $(\mu - 1)t/\mu$ from the first. The one nodal point is of course the image of the other, and the lens in this case acts simply as a plane parallel slab of glass with a point object placed close up to one of its faces.

In the case of a meniscus lens or an unsymmetrical double-convex or double-concave lens the position of the optical centre of the lens can easily be found graphically, and the equivalent planes can be shown in the same way to divide the distances from it to the surfaces of the lens in the ratio $\mu - 1$ to 1, being of course nearer the optical centre.

The positions of the equivalent planes in some typical cases are shown in fig. 49.

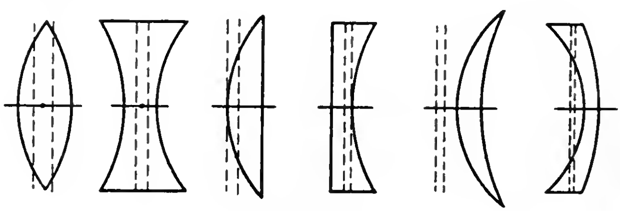


FIG. 49.

§ To find the equivalent focal length of a thick lens we set out from equations (5) and (6) of p. 26. They are

$$\frac{\mu}{s} - \frac{1}{u} = \frac{\mu - 1}{r_1},$$

$$\frac{1}{v} - \frac{\mu}{s + t} = \frac{1 - \mu}{r_2},$$

where u is the distance of the object from the first face, s the distance of the first image from the first face, and v the distance of the final image from the second face. Previously t was neglected in comparison with s, but this simplification is now no longer admissible.

The equations can be written :—

$$s = \frac{\mu}{\frac{\mu - 1}{r_1} + \frac{1}{u}},$$

$$s + t = \frac{\mu}{\frac{1}{v} - \frac{1 - \mu}{r_2}}.$$

Hence on eliminating s we obtain

$$t = \frac{\mu}{\frac{1}{v} - \frac{1-\mu}{r_2}} - \frac{\mu}{\frac{\mu - 1}{r_1} + \frac{1}{u}},$$

which gives

$$\frac{t}{\mu}\left\{1 - v\frac{(1-\mu)}{r_2}\right\}\left\{1 + u\frac{(\mu-1)}{r_1}\right\} = v\left\{1 + u\frac{(\mu-1)}{r_1}\right\} - u\left\{1 - v\frac{(1-\mu)}{r_2}\right\}$$

or

$$uv\left\{\frac{t}{\mu}\frac{(\mu-1)^2}{r_1 r_2} - \frac{(\mu-1)}{r_1} + \frac{(\mu-1)}{r_2}\right\} + v\left\{\frac{t}{\mu}\frac{(\mu-1)}{r_2} - 1\right\} + u\left\{\frac{t}{\mu}\frac{(\mu-1)}{r_1} + 1\right\} + \frac{t}{\mu} = 0.$$

Denote the coefficients for convenience by A, B, C, D; then the equation may be written

$$\mathrm{A}uv + \mathrm{B}u + \mathrm{C}v + \mathrm{D} = 0 \quad . \quad . \quad . \quad (18)$$

It may easily be shown that $\mathrm{BC} - \mathrm{AD} = -1$. Equation (18) can be written

$$uv + \frac{\mathrm{B}}{\mathrm{A}}u + \frac{\mathrm{C}}{\mathrm{A}}v + \frac{\mathrm{BC}}{\mathrm{A}^2} = \frac{\mathrm{BC} - \mathrm{AD}}{\mathrm{A}^2},$$

or

$$\left(u + \frac{\mathrm{C}}{\mathrm{A}}\right)\left(v + \frac{\mathrm{B}}{\mathrm{A}}\right) = -\frac{1}{\mathrm{A}^2} \quad . \quad . \quad . \quad (19)$$

By comparison with (12) $u + \frac{\mathrm{C}}{\mathrm{A}} = 0$ and $v + \frac{\mathrm{B}}{\mathrm{A}} = 0$ obviously give the focal planes of the lens and the equivalent focal length is given by

$$\frac{1}{f} = -\mathrm{A} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2} - \frac{(\mu-1)t}{\mu r_1 r_2}\right).$$

There is no ambiguity in the sign since, when $t = 0$, f must reduce to its value for a thin lens. Note that in equation (19) u and v are not measured from the same point.

The equivalent focal length of a thick plano-convex or plano-concave lens can be found very easily without having recourse to the above formula. For consider fig. 50. Let DA be an incident ray

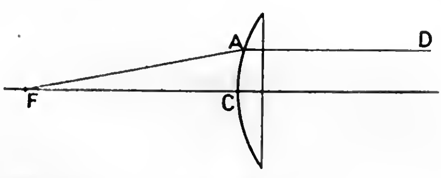


FIG. 50.

parallel to the axis. It passes through the plane surface without suffering any deviation. To apply the formula

$$\frac{\mu_2}{v} - \frac{\mu_1}{u} = \frac{\mu_2 - \mu_1}{r}$$

to the second surface we must write $u = \infty$, $\mu_2 = 1$ and $\mu_1 = \mu$. This gives

$$v = \frac{r}{1 - \mu}.$$

Hence CF $= r/(1 - \mu)$, and since C is the nodal point of the image space, CF is the focal length of the lens.

Spherical Lens. In the case of a spherical lens surrounded by air it is clear that the two nodal points coincide at the centre of the sphere, and that consequently the two principal planes coincide with the diametral plane there.

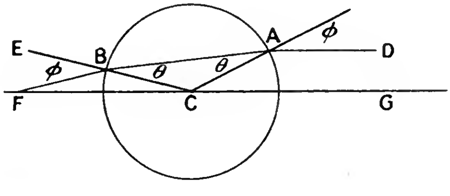


FIG. 51.

To find the equivalent focal length consider a ray DA entering the lens at A, leaving it at B, and intersecting the axis at F. Let ϕ be the angle of incidence and θ the angle of refraction at A, let r be the radius of the sphere and μ the index of refraction of the material of which it is composed. Then $\angle$ABC $= \theta$ and $\angle$EBF $= \phi$; also since AC is inclined at a small angle to the axis, angles ϕ and θ are small, and instead of $\sin\phi = \mu \sin\theta$ we may write $\phi = \mu\theta$. $\angle$ACG $= \phi$; consequently $\angle$BCF $= \pi - \angle$BCA $- \phi = \pi - (\pi - 2\theta) - \phi = 2\theta - \phi$ and hence $\angle$BFC $= 2(\phi - \theta)$.

In $\triangle$FBC

$$\frac{\text{FC}}{\text{BC}} = \frac{\sin \text{FBC}}{\sin \text{BFC}} = \frac{\sin\phi}{\sin 2(\phi - \theta)} = \frac{\phi}{2(\phi - \theta)},$$

since the angles are small. Hence the equivalent focal length FC

$$= \text{BC}\frac{\phi}{2(\phi - \theta)} = \frac{\mu r}{2(\mu - 1)}.$$

Cardinal Points of a Spherical Refracting Surface. Let the medium left of the spherical surface have the index of refraction μ and let the medium right of the surface have the index of refraction 1. Then it is clear both nodal points coincide in N, the centre of curvature of the surface. As in this case the initial and final media are not the same, the principal planes do not pass through the nodal points.

From the formula

$$\frac{\mu}{v} - \frac{1}{u} = \frac{\mu - 1}{r}$$

by making in succession u and v infinite, we find that the focal planes of the image and object spaces are at F_2 and F_1 where $AF_2 = \mu r/(\mu - 1)$ and $AF_1 = -r/(\mu - 1)$. The principal planes coincide and pass through A.

In the case of a spherical mirror the nodal points coincide with the centre of curvature of the mirror, the two principal planes with its

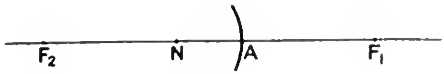


FIG. 52.

surface and the focal planes with its focus. In the case of a mirror BC − AD is positive in equation (12), and f_1 and f_2 as defined in (14) are equal and have the same sign.

Two Thin Lenses. In the case of two thin lenses on the same axis separated by a distance d the focal planes, the equivalent focal length, and the nodal points can all be found very easily by graphical construction.

Let ϕ_1 be the focal length of the first lens and ϕ_2 the focal length of the second. Let LB (fig. 53) be a ray incident on the first lens

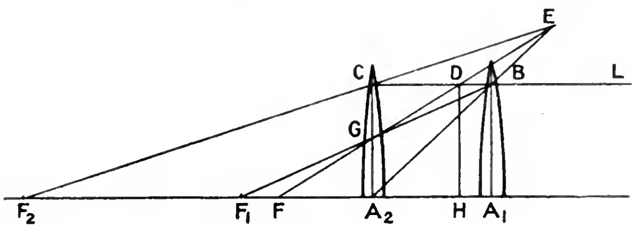


FIG. 53.

parallel to the axis. Then after refraction it passes through F_1, the focus of the first lens.

Now find the image of B produced by the second lens in the usual manner, that is by drawing two rays through it, one parallel to the axis which passes after refraction through F_2, the focus of the second lens and another BA_2 through the centre of the second lens. These two rays meet in E; therefore E is the image of B produced by the second lens. Consequently EG gives the direction of BG after refraction by the second lens and the point F, where EG meets the axis, is the focus of the image space.

Let EG cut BC in D and draw DH perpendicular to the axis to meet it in H. The plane DH is then the equivalent plane of the image space, since the line DH obviously gives the distance from the axis at which LB meets the equivalent plane of the object space. HF is consequently the equivalent focal length of the system.

Since the lines CA_2, DF, and BF_1 meet in a point and CB and F_1A_2 are parallel

$$\frac{CD}{FA_2} = \frac{CB}{F_1A_2} = \frac{d}{-\phi_1 - d} \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Similarly, since CF_2, DF, and BA_2 meet in a point

$$\frac{DB}{FA_2} = \frac{CB}{F_2A_2} = \frac{d}{-\phi_2} \quad . \quad . \quad . \quad . \quad . \quad (21)$$

Hence adding (20) and (21) we obtain

$$\frac{CD}{FA_2} + \frac{DB}{FA_2} = \frac{d}{-\phi_1 - d} + \frac{d}{-\phi_2}$$

or

$$\frac{CB}{FA_2} = \frac{-d(\phi_1 + \phi_2 + d)}{(\phi_1 + d)\phi_2},$$

which gives

$$FA_2 = -\frac{(\phi_1 + d)\phi_2}{\phi_1 + \phi_2 + d}.$$

The equivalent focal length of the system is then given by

$$HF = HA_2 + A_2F = DC + A_2F = A_2F(1 + DC/A_2F)$$

$$= \frac{(\phi_1 + d)\phi_2}{\phi_1 + \phi_2 + d}\left(1 + \frac{d}{-\phi_1 - d}\right) \text{ from (20)}$$

$$= \frac{\phi_1\phi_2}{\phi_1 + \phi_2 + d}.$$

The nodal points can be found more directly as follows. Let KJ be a focal plane of the lens A_1B_1 and MG a focal plane of the lens

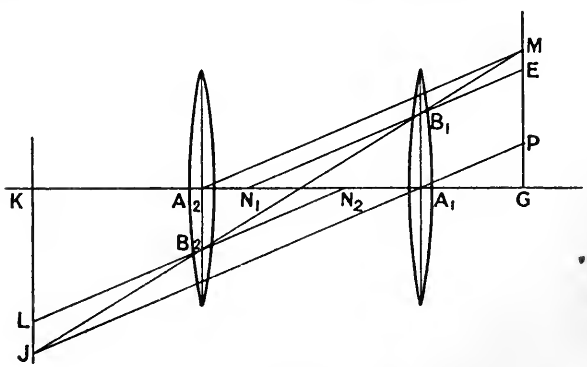


Fig. 54.

A_2B_2. Through the centres of the lenses draw any two parallel rays A_1J and A_2M to meet these planes in J and M. Join JM and let it cut the two lenses in B_1 and B_2. Through B_1 and B_2 draw B_1N_1 and B_2N_2 parallel to the first two rays to meet the axis in N_1 and N_2. Then the ray EB_1 after refraction by the first lens must take the direction B_1B_2 since it meets the ray PA_1 in J. Similarly the ray B_1B_2 must take the direction B_2L after refraction by the second lens. Consequently EB_1B_2L is the course of a ray through the system, N_1 is the nodal point of the object space, and N_2 the nodal point of the image space.

Now $$\frac{A_1N_1}{A_1A_2} = \frac{PE}{PM} = \frac{JB_1}{JM} = \frac{KA_1}{KG} = \frac{\phi_1}{\phi_1 + \phi_2 + d}.$$

Hence $$A_1N_1 = -\frac{\phi_1 d}{\phi_1 + \phi_2 + d}.$$

Similarly $$A_2N_2 = \frac{\phi_2 d}{\phi_1 + \phi_2 + d}.$$

Fig. 55 shows the variation in the position of the equivalent planes and focal planes of a system of two thin convex lenses as the distance between them is gradually increased. In it ϕ_1 and ϕ_2 are

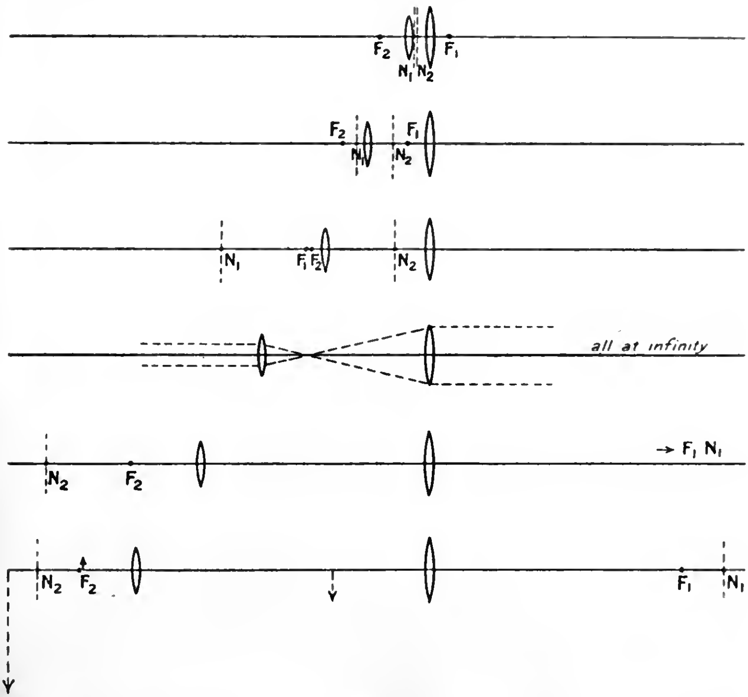


FIG. 55.

taken respectively as -3 and -1, and d is given in succession the values of ·5, 1·5, 2·5, 4·0, 5·5, and 7·0. The lens, the focal length of which is denoted by ϕ_2, is on the left.

It will be noticed that in the first position the equivalent planes are crossed and that, as the separation of the lenses increases, the equivalent planes move further apart, while at the same time the equivalent focal length becomes greater. In the second and third figures the focus of the object space is virtual.

In the fourth case, which represents the telescope, the lens on the right being the object glass, the cardinal points are all at infinity, and parallel light entering the system always emerges as parallel light.

In the fifth and sixth cases the character of the system has changed, the equivalent planes reappearing at infinity on the sides opposite to those on which they disappeared, while the focal length has changed sign. The sixth case represents the compound microscope, the lens on the left being the object glass. In order to make the connection with the telescope and the microscope evident, in the fourth case the paths of the rays and in the sixth case the images are shown by dotted lines.

Telescopic Systems. If A is zero in the fundamental equation (12) we obtain

$$\mathrm{B}x_1 + \mathrm{C}x_2 + \mathrm{D} = 0$$

and find that if x_1 is infinite x_2 is infinite, and vice versa. Consequently light which enters the system parallel emerges parallel. In this case the system is said to be telescopic. It is exemplified by the fourth case in fig. 55.

EXAMPLES.

(1) Prove Helmholtz's law of magnification for the case of reflection at a concave or convex spherical surface.

(2) An object is displaced a small distance *du* along the axis of a thin lens. Find an expression for the corresponding displacement *dv* of the image. If *du* is the length of an object placed along the axis, *dv* is the length of the corresponding image and the ratio *dv*/*du* may be referred to as the longitudinal magnification. Show that it is equal to the square of the ordinary lateral magnification.

(3) Prove that if an object be at such a distance from a thick convex lens that it forms an image of equal size at the other side, the distance from object to image, minus the distance between the two principal points, is equal to four times the focal length.

(4) On one side of a bi-convex thin lens the medium is water; on the other side it is air. The radii of curvature of the two faces of the lens are each equal to 20 cms. and it is made of glass of index of refraction 1·52. Find the positions of the focal planes, principal planes, and nodal points.

(5) A glass sphere has an index of refraction 1·5 and radius of curvature 2 cms. Where will it form an image of an object distant 5 cms. from the centre of the sphere, and what will be the magnification of this image?

(6) On one side of a spherical glass lens of radius 1 cm. and index of refraction 1·52 the medium is air and on the other side it is water. Find the positions of the focal planes, principal planes, and nodal points, and use them to find the positions and magnifications of the image, when the object is situated in air (*a*) 4 cms., (*b*) 1·5 cms. from the centre of the sphere.

(7) Solve the second part of the preceding question by applying the formula for refraction at a spherical surface to each of the surfaces of the lens in succession.

(8) Show that if f is the focal length of a lens combination capable of giving a real image, and if the lens combination is placed so as to give an image of an object on a screen 1 metre distant from the principal plane of the image space, then the magnification of the image will be $100/f - 1$.

(9) A glass hemisphere of radius r and refractive index μ is treated as a lens, rays passing through it being limited to those nearly coinciding with the axis. Show that one principal point coincides with the intersection of the convex surface with the axis, while the other principal point is within the lens at a distance r/μ from the plane surface. Prove also that the focal length of the lens is equal to $r/(1 - \mu)$.

(10) A plano-convex lens of glass of index of refraction 1·52 and radius of curvature 24 cms. is 2 cms. thick measured along the axis. Calculate its focal length, and find the position of the image when the object is distant 50 cms. from the convex surface (a) on the convex side, (b) on the plane side.

(11) If in the case of the lens described in the previous question the medium on the curved face is water and on the other face air, find the situation of the cardinal points, and calculate the positions of the image when the object is situated in air 50 cms. distant from the plane face.

(12) What does the formula for the equivalent focal length of a system of two coaxial thin lenses become when the space between them is filled with water?

(13) Two similar plano-convex lenses are placed with their plane faces together and then drawn apart to a short distance. Show that, when separated, the combination has a greater focal length than when they are in contact. Show also that, when separated, the positions of the principal foci are nearer to the respective curved surfaces than when the lenses were in contact.

(14) An object is placed on the axis of a concave mirror beyond the focus and a plate of glass of thickness t and refractive index μ interposed between the focus and the mirror, the axis of the mirror being normal to the plate. Show that the effect on the position of the image is the same as if the mirror had been displaced through a distance $t(\mu - 1)/\mu$ towards the object.

CHAPTER IV.

THE DEFECTS OF THE IMAGE.

In Chapter II we assumed that the rays falling on the mirrors and lenses lay near the axis and were only slightly inclined to the latter. In this case a point image is formed of a point object. It is now necessary to drop the restriction and inquire what happens when the rays are inclined to the axis at a considerable angle.

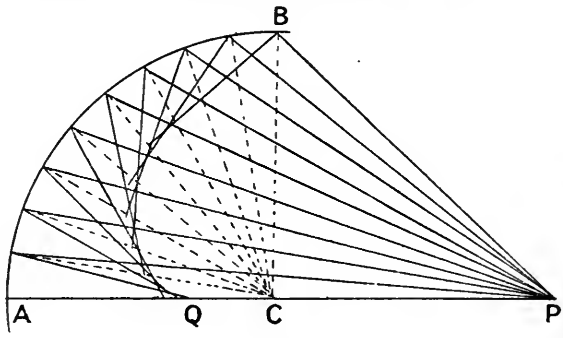


Fig. 56.

Let AB represent a section of a concave spherical mirror, let C be its centre of curvature, and let P be a point object. Then if rays be drawn diverging from P at all angles, by making their angles of reflection equal to their angles of incidence they may be found graphically to occupy after reflection the positions shown in the figure.

Only the rays reflected from the neighbourhood of A pass through the image of P at Q; the others intersect the axis between A and Q. Any two rays reflected from neighbouring points of the mirror intersect each other before reaching the axis and these points of intersection lie on a curve termed the caustic curve. All the reflected rays touch this curve and it has a cusp at Q. The form of the caustic alters, of course, if P moves along the axis. Owing to the reflected rays coming closer together at the caustic, a bright curve is formed on a piece of paper or other white surface there. A familiar example of this is nearly horizontal sunlight shining into a teacup filled almost to the top with milk. Here the inside of the cup acts as mirror. A better way of producing the caustic is by means of a piece of polished steel spring bent circular and placed on a drawing-board.

If fig. 56 be rotated about AP as axis, instead of a plane pencil of rays we have a solid conical pencil diverging from P and the caustic curve traces out a surface called the caustic surface. Also all the rays which diverge from P at the same angle with the axis pass after reflection through the same point on the axis. There will consequently be a bright line on the axis from A to Q, and this line may be regarded as part of the caustic surface. The experiment with the teacup is apt to produce a false impression about the caustic produced by a spherical mirror. The teacup being cylindrical does not produce the bright line AQ.

In fig. 57 DD′ represents a section of a smaller portion of the same spherical mirror. The rays DG and D′G from the margin of the

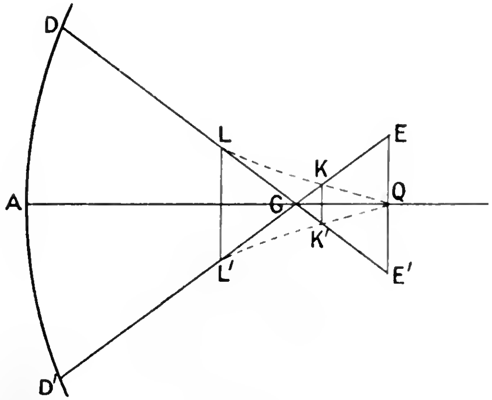


Fig. 57.

mirror intersect in G while Q, as before, is the image formed by rays inclined at a small angle to the axis. The caustic is shown by the dotted curve.

If a screen were placed at LL′ to receive the light, it would be illuminated by a circular patch with a bright edge. If it were moved towards G the patch would contract; at G a bright spot would appear in the centre. At KK′ the patch would have reached its smallest diameter; thereafter a dimmer zone would appear on the outside and the whole patch would increase in size while its central bright part would contract. Finally at Q we would have a brilliant point of light surrounded by a dim circular disc.

The circle KK′ is called the circle of least confusion and may be regarded as the nearest approach to an image of the luminous point formed by the mirror. The distance GQ is called the longitudinal spherical aberration, or simply the aberration of the marginal ray DG, while QE is called its lateral spherical aberration.

4 *

Spherical Aberration of a Concave Mirror. To find the magnitude of the aberration consider fig. 58 in which P is the object, C the centre

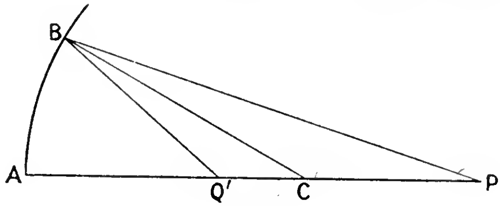


FIG. 58.

of curvature, and Q′ the point in which a ray diverging at a wide angle from the axis meets it after reflection. Let AP $= u$, AC $= r$, and AQ′ $= v'$. Also let AB $= h$, a quantity such that the cube and higher powers of h/r may be neglected. It is immaterial therefore whether we measure h along the arc or take it to represent the perpendicular distance of B from the axis.

In $\triangle$BCP $$\frac{\text{CP}}{\text{PB}} = \frac{\sin \text{CBP}}{\sin \text{PCB}}.$$

In $\triangle$BQ′C $$\frac{\text{Q'C}}{\text{BQ'}} = \frac{\sin \text{Q'BC}}{\sin \text{BCQ'}}.$$

But $\angle$CBP $= \angle$Q′BC and sin PCB $=$ sin BCQ′.

Therefore $$\frac{\text{CP}}{\text{PB}} = \frac{\text{Q'C}}{\text{BQ'}} \quad . \quad . \quad (22)$$

Now $$\text{PB}^2 = \text{CP}^2 + \text{CB}^2 + 2\text{CP}\,.\,\text{CB} \cos \text{ACB}$$
$$= (u - r)^2 + r^2 + 2r(u - r) \cos \frac{h}{r}$$

and $\cos \dfrac{h}{r} = 1 - \dfrac{h^2}{2r^2}$ to the order of approximation adopted.

Hence $$\text{PB}^2 = u^2 - \frac{(u - r)}{r} h^2$$

and $$\text{PB} = u\left\{1 - \left(\frac{1}{r} - \frac{1}{u}\right)\frac{h^2}{2u}\right\} \text{ approximately.}$$

Similarly $$\text{Q'B} = v'\left\{1 - \left(\frac{1}{r} - \frac{1}{v'}\right)\frac{h^2}{2v'}\right\} \text{ approximately.}$$

From (22) CP . BQ′ $=$ Q′C . PB. This gives on substitution

$$(u - r)\,v'\left\{1 - \left(\frac{1}{r} - \frac{1}{v'}\right)\frac{h^2}{2v'}\right\} = (r - v')\,u\left\{1 - \left(\frac{1}{r} - \frac{1}{u}\right)\frac{h^2}{2u}\right\}$$

or $$\left(\frac{1}{r} - \frac{1}{u}\right)\left\{1 - \left(\frac{1}{r} - \frac{1}{v'}\right)\frac{h^2}{2v'}\right\} = \left(\frac{1}{v'} - \frac{1}{r}\right)\left\{1 - \left(\frac{1}{r} - \frac{1}{u}\right)\frac{h^2}{2u}\right\},$$

if each side be divided by $uv'r$. Hence

$$\frac{1}{v'} + \frac{1}{u} = \frac{2}{r} + \left(\frac{1}{r} - \frac{1}{u}\right)\left(\frac{1}{v'} - \frac{1}{r}\right)\left(\frac{1}{u} + \frac{1}{v'}\right)\frac{h^2}{2} \quad . \quad . \quad (23)$$

Since h^2 is small, we can substitute v for v' in its coefficient, where v denotes the image distance for rays inclined at a small angle to the axis. Then, since $\frac{1}{v} + \frac{1}{u} = \frac{2}{r}$, equation (23) becomes

$$\frac{1}{v'} + \frac{1}{u} = \frac{2}{r} + \left(\frac{1}{r} - \frac{1}{u}\right)^2 \frac{h^2}{r} \quad . \quad . \quad . \quad (24)$$

This equation gives the position of Q′.

The aberration of the ray BQ′, $v' - v$, is equal to

$$\left(\frac{1}{v} - \frac{1}{v'}\right) vv'$$

$$= - \left(\frac{1}{r} - \frac{1}{u}\right)^2 \frac{h^2}{r} vv'$$

$$= - \frac{\left(\frac{1}{r} - \frac{1}{u}\right)^2}{\left(\frac{2}{r} - \frac{1}{u}\right)^2} \frac{h^2}{r}$$

on putting $v = v'$.

Astigmatic Reflection at a Concave Mirror. Let BD be a section of a small portion of a concave spherical mirror of centre of curvature C, and let rays PB and PD from a luminous point P be incident on the

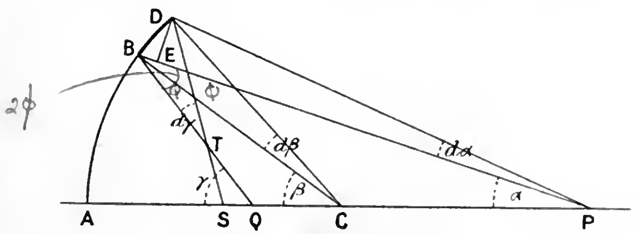


FIG. 59.

mirror at a large angle with the axis. After reflection let the ray PB meet the axis in Q, and let the ray PD meet BQ in T and the axis in S. Denote angles BPC, BCQ, and BQA by α, β, and γ, and let angles DPB, DCB, and DTB be $d\alpha$, $d\beta$, and $d\gamma$. Let BP = u, BT = v_1, BQ = v_2, and CB = r, and denote the angles of incidence and reflection at B by ϕ. Then

$$\triangle \mathrm{QBP} = \triangle \mathrm{QBC} + \triangle \mathrm{CBP},$$

which gives

$$v_2 u \sin 2\phi = v_2 r \sin \phi + ru \sin \phi$$

on omitting the common factor $\frac{1}{2}$. Divide throughout by $uv_2 r \sin \phi$ and we obtain

$$\frac{1}{u} + \frac{1}{v_2} = \frac{2 \cos \phi}{r} \quad . \quad . \quad . \quad . \quad . \quad (25)$$

From D draw DE perpendicular to BP. Then since $d\alpha$ is small,

$DE = DP\,d\alpha = u\,d\alpha$. Since BD is small we may regard it as straight. $\angle BDE = \phi$. Consequently $DE = BD \cos \phi$. Thus $u\,d\alpha = BD \cos \phi$

or
$$d\alpha = \frac{BD \cos \phi}{u}.$$

Similarly
$$d\gamma = \frac{BD \cos \phi}{v_1},$$

while
$$d\beta = \frac{BD}{r}.$$

Now $\phi = \beta - \alpha = \gamma - \beta$. Hence $\alpha + \gamma = 2\beta$. This relation must hold also for the increments of α, β, and γ; thus

$$d\alpha + d\gamma = 2d\beta,$$

or
$$\frac{BD \cos \phi}{u} + \frac{BD \cos \phi}{v_1} = \frac{2BD}{r}.$$

On dividing by $BD \cos \phi$ we thus obtain

$$\frac{1}{u} + \frac{1}{v_1} = \frac{2}{r \cos \phi} \qquad . \quad . \quad . \quad . \quad (26)$$

Now rotate the diagram through a small angle about AP. The results hold for every instantaneous position. BD sweeps out an element of area, approximately rectangular in shape. The triangle PBD sweeps out a solid pencil of rays diverging from the point P. The point T traces out a short line, and the reflected rays pass through this line and the axis. This is shown more effectively in fig. 60 where BDD′B′ is the element of area traced out by BD, TT′ the line traced out by the point T, and SQ the axis.

Thus a pencil of rays, which diverges from a point, after oblique reflection at an element of a concave mirror converges to two short lines, one perpendicular to the plane containing the principal ray of the pencil and the centre of curvature of the mirror and the other in this plane. These lines are known as the focal lines, TT′ being called the first focal line and SQ the second focal line, and their positions are given by equations (26) and (25).

If a screen is placed in the way of the pencil near a focal line, an irregular patch of light is produced, either the length or the breadth of which is extremely small. But at a certain place between the focal lines the length and the breadth of the patch are equal. This place is known as the circle of least confusion, because in the general case the patch is approximately circular there. It may be regarded as a blurred image of the point P, the nearest approach to an image formed by the reflected pencil.

A pencil such as that represented in fig. 60, which does not anywhere pass through a point, is called an astigmatic pencil from the Greek words *a* not and *stigma* a point. The distance between the focal lines is called the astigmatic difference. It is given by

$$v_2 - v_1 = v_1 v_2 \left(\frac{1}{v_1} - \frac{1}{v_2}\right)$$
$$= v_1 v_2 \left(\frac{2}{r \cos \phi} - \frac{2 \cos \phi}{r}\right)$$
$$= \frac{2v_1 v_2}{r} \sin \phi \tan \phi.$$

It thus increases very rapidly with the angle of incidence.

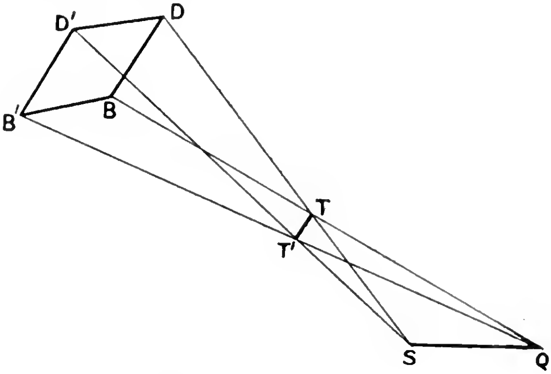


FIG. 60.

Curvature and Distortion. So far we have been concerned with point objects. Fig. 61 represents a line object at P, and if the positions

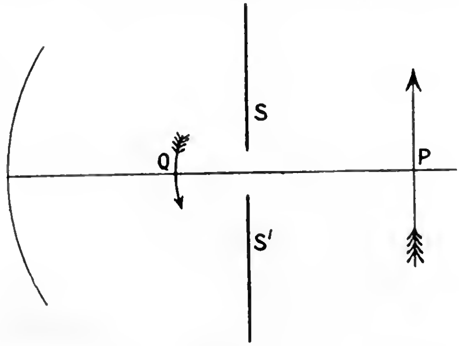


FIG. 61.

of the images of every point on it formed by the concave mirror in the diagram are calculated by the elementary formula, they are found to lie along the curved inverted arrow at Q. If a screen SS′ with a small aperture is placed as shown at the centre of curvature of the mirror, then the image of every point in P is formed only by a thin centric pencil, and the defects due to spherical aberration and astigmatism are entirely eliminated at the expense of the brightness of the image. This is an interesting example of the influence of a stop at the right place. Owing to all the points in the object not being at the same distance from the mirror, as measured along the straight line through the centre of curvature, the magnification of the ends of the object is different from the magnification of its middle.

In this case, therefore, we have two new defects, curvature and distortion of the image.

The caustic formed by reflection at a convex spherical mirror, the aberration of the marginal ray and the astigmatic reflection of a thin pencil at a convex mirror can all be treated in the same way as in the corresponding case for the concave mirror.

Refraction of a Wide Angle Pencil at a Spherical Surface. Just as in the analogous case of reflection, if a wide angle pencil of rays diverges from a point and is refracted at a spherical surface, after refraction the rays do not pass through a point but touch a caustic surface which has a cusp on the axis, and we have the same phenomena of aberration and the circle of least confusion.

To find the position in which the marginal ray meets the axis after refraction consider fig. 62. PB is a ray incident from air on the

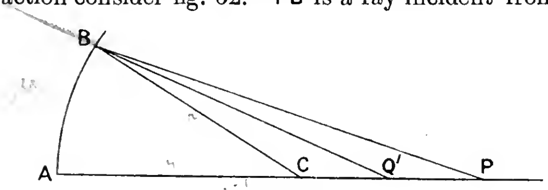


FIG. 62.

spherical surface AB of a medium of index of refraction μ. C is the centre of curvature of the surface, BQ′ is the path of the ray after refraction produced back, and h denotes the distance AB. As on p. 52, h is a quantity such that the cube and higher powers of h/r may be neglected. Let AP $= u$, AQ′ $= v'$, and AC $= r$.

Then in $\triangle$BCP $$\frac{CP}{PB} = \frac{\sin CBP}{\sin PCB}$$

and in $\triangle$BCQ′ $$\frac{CQ'}{Q'B} = \frac{\sin CBQ'}{\sin Q'CB}.$$

But $$\sin CBP = \mu \sin CBQ'. \quad \text{Hence} \quad \frac{CP}{PB} = \mu\frac{CQ'}{Q'B}$$

or $$\mu CQ' . PB = CP . Q'B \quad . \quad . \quad . \quad . \quad (27)$$

Now $$BP^2 = BC^2 + CP^2 + 2BC . CP \cos ACB$$

$$= r^2 + (u - r)^2 + 2r(u - r) \cos\frac{h}{r}$$

$$= r^2 + (u - r)^2 + 2r(u - r)\left(1 - \frac{h^2}{2r^2}\right)$$ to the order of approximation involved

$$= u^2 - (u - r)\frac{h^2}{r}.$$

Hence, remembering that h^2 is small, we obtain

$$\text{BP} = u\left\{1 - \left(\frac{1}{r} - \frac{1}{u}\right)\frac{h^2}{2u}\right\}$$

and similarly

$$\text{BQ}' = v'\left\{1 - \left(\frac{1}{r} - \frac{1}{v'}\right)\frac{h^2}{2v'}\right\}.$$

Substitution in (27) gives

$$\mu(v' - r)u\left\{1 - \left(\frac{1}{r} - \frac{1}{u}\right)\frac{h^2}{2u}\right\} = (u - r)v'\left\{1 - \left(\frac{1}{r} - \frac{1}{v'}\right)\frac{h^2}{2v'}\right\}$$

or

$$\mu\left(\frac{1}{r} - \frac{1}{v'}\right)\left\{1 - \left(\frac{1}{r} - \frac{1}{u}\right)\frac{h^2}{2u}\right\} = \left(\frac{1}{r} - \frac{1}{u}\right)\left\{1 - \left(\frac{1}{r} - \frac{1}{v'}\right)\frac{h^2}{2v'}\right\}$$

or

$$\frac{\mu}{v'} - \frac{1}{u} = \frac{\mu - 1}{r} + \left(\frac{1}{r} - \frac{1}{v'}\right)\left(\frac{1}{r} - \frac{1}{u}\right)\left(\frac{1}{v'} - \frac{\mu}{u}\right)\frac{h^2}{2}.$$

Since h^2 is small, it is permissible to substitute v for v' in the coefficient of h^2, where v is the image distance for rays inclined at a small angle to the axis given by

$$\frac{\mu}{v} - \frac{1}{u} = \frac{\mu - 1}{r}.$$

Hence the equation determining the position of Q′ becomes

$$\frac{\mu}{v'} - \frac{1}{u} = \frac{\mu - 1}{r} + \frac{1}{\mu^2}\left(\frac{1}{r} - \frac{1}{u}\right)^2\left(\frac{1}{u} + \frac{\mu - 1}{r} - \frac{\mu^2}{u}\right)\frac{h^2}{2}$$

$$= \frac{\mu - 1}{r} + \frac{\mu - 1}{\mu^2}\left(\frac{1}{r} - \frac{1}{u}\right)^2\left(\frac{1}{r} - \frac{\mu + 1}{u}\right)\frac{h^2}{2} \quad . \quad (28)$$

Spherical Aberration of a Thin Lens. Let us apply equation (28) to both sides of a lens of negligible thickness. In conformity with the notation of Chapter II let u be the distance of the object, r_1, r_2 the radii of curvature of the first and second faces, μ the index of refraction of the material of the lens, s the distance of the point where the direction of the marginal ray cuts the axis after the first refraction, and v' the corresponding distance after the second refraction.

Then for the first refraction

$$\frac{\mu}{s} - \frac{1}{u} = \frac{\mu - 1}{r_1} + \frac{\mu - 1}{\mu^2}\left(\frac{1}{r_1} - \frac{1}{u}\right)^2\left(\frac{1}{r_1} - \frac{\mu + 1}{u}\right)\frac{h^2}{2}.$$

For the second refraction, on the supposition that the direction of the ray is reversed,

$$\frac{\mu}{s} - \frac{1}{v'} = \frac{\mu - 1}{r_2} + \frac{\mu - 1}{\mu^2}\left(\frac{1}{r_2} - \frac{1}{v'}\right)^2\left(\frac{1}{r_2} - \frac{\mu + 1}{v'}\right)\frac{h^2}{2}.$$

By subtraction

$$\frac{1}{v'} - \frac{1}{u} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) + \frac{\mu - 1}{\mu^2}\left\{\left(\frac{1}{r_1} - \frac{1}{u}\right)^2\left(\frac{1}{r_1} - \frac{\mu + 1}{u}\right) - \left(\frac{1}{r_2} - \frac{1}{v'}\right)^2\left(\frac{1}{r_2} - \frac{\mu + 1}{v'}\right)\right\}\frac{h^2}{2}.$$

In the coefficient of h^2 instead of v' it is permissible to use v, the value for axial rays, which is given by the equation

$$\frac{1}{v} - \frac{1}{u} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right).$$

Hence

$$\frac{1}{v'} - \frac{1}{u} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) + \frac{\mu - 1}{\mu^2}\left\{\left(\frac{1}{r_1} - \frac{1}{u}\right)^2\left(\frac{1}{r_1} - \frac{\mu + 1}{u}\right) - \left(\frac{1}{r_2} - \frac{1}{v}\right)^2\left(\frac{1}{r_2} - \frac{\mu + 1}{v}\right)\right\}\frac{h^2}{2}.$$

The aberration of the marginal ray, $v' - v$,

$$= \left(\frac{1}{v} - \frac{1}{v'}\right)vv' = -\frac{\mu - 1}{\mu^2}\left\{\left(\frac{1}{r_1} - \frac{1}{u}\right)^2\left(\frac{1}{r_1} - \frac{\mu + 1}{u}\right) - \left(\frac{1}{r_2} - \frac{1}{v}\right)^2\left(\frac{1}{r_2} - \frac{\mu + 1}{v}\right)\right\}\frac{v^2h^2}{2} \quad . \quad . \quad . \quad (29)$$

Spherical Aberration of a Thin Lens when the Object is at Infinity. Write $u = \infty$, $v = f$ in (29). Then the expression becomes

$$-\frac{\mu - 1}{\mu^2}\left\{\frac{1}{r_1^3} - \left(\frac{1}{r_2} - \frac{1}{f}\right)^2\left(\frac{1}{r_2} - \frac{\mu + 1}{f}\right)\right\}\frac{f^2h^2}{2} \quad . \quad . \quad (30)$$

Put $\sigma = r_1/r_2$; then by means of the equations

$$\frac{1}{f} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) = \frac{(\mu - 1)(1 - \sigma)}{r_1} = \frac{(\mu - 1)(1 - \sigma)}{\sigma r_2} \quad . \quad (31)$$

eliminate r_1 and r_2 from (30). The result is

$$\frac{-h^2\{2 - 2\mu^2 + \mu^3 + \sigma(\mu + 2\mu^2 - 2\mu^3) + \sigma^2\mu^3\}}{f2\mu(\mu - 1)^2(1 - \sigma)^2} \quad . \quad (32)$$

If the lens is double-convex or double-concave σ is negative; if it is convexo-concave or concavo-convex σ is positive.

If the expression within the bracket in the numerator is regarded as a quadratic function of σ, the discriminant is

$$(\mu + 2\mu^2 - 2\mu^3)^2 - 4\mu^3(2 - 2\mu^2 + \mu^3)$$
$$= \mu^2(1 - 4\mu).$$

As μ varies in practice only between 1·5 and 2, this is always negative. The zeroes of the function are consequently imaginary and it does not change sign as σ varies. But it is positive when $\sigma = 0$; hence it is always positive and $v' - v$ has consequently the same sign as $-f$. Thus the marginal rays always come to a focus nearer the lens than the axial rays do, no matter what the ratio of the radii of curvature is. The aberration can never be made zero in the case of a single thin lens, but a convex and concave lens can be combined so that the aberration of the combination is zero.

Let us suppose that h, f, and μ are given, and that it is required to find for what value of σ the aberration is a minimum. Differentiating (32) with reference to σ and equating the result to zero we find

$$\sigma = \frac{2\mu^2 - \mu - 4}{\mu(1 + 2\mu)} \quad . \quad . \quad . \quad . \quad (33)$$

This obviously gives a minimum since it is the only turning value and the expression is infinite for $\sigma = 1$. If μ be put equal to 1·5 in (33), σ becomes $-\frac{1}{6}$, and if μ be put equal to 2, σ is $+\frac{1}{5}$. In both cases, according to (31), r_1 must have the same sign as f. Consequently, if the lens is to be convex, in the first case it must be double-convex with the light incident first on the more curved face, and in the second case it must be a convex meniscus with the light incident first on the more curved face. A lens, the ratio of the curvatures of which is chosen so as to make the spherical aberration a minimum, is termed a crossed lens.

The following figures taken from Drude's "Optics" show how the longitudinal aberration varies with μ and σ, when f and h, the focal length and radius of the lens, are kept at the constant values of 1 metre and 10 centimetres.

Shape of Lens.	$\mu = 1·5$.		$\mu = 2·0$.	
	σ.	Aberration.	σ.	Aberration.
Front surface plane .	∞	4·5 cms.	∞	2·0 cms.
Symmetrical . .	-1	1·67 ,,	-1	1·0 ,,
Back surface plane .	0	1·16 ,,	0	·50 ,,
Most favourable form	$-\frac{1}{6}$	1·07 ,,	$+\frac{1}{5}$	·44 ,,

It will be noticed from the table that the crossed lens has not much advantage over the plano-convex lens, and also that increasing the index of refraction diminishes the aberration considerably.

Reversing the plano-convex lens has a great effect on its spherical aberration. This is to be expected from elementary considerations. When the parallel rays fall first on the plane face, all the refraction takes place at the curved face; when they fall first on the curved face, the refraction is distributed between the two faces and it is not necessary for the angle of incidence to be so great. The elementary theory which ignores aberration assumes $\phi = \mu\theta$ for $\sin \phi = \mu \sin \theta$. The theory given here is equivalent to taking in the next term in the expansion and assuming $\left(\phi - \frac{\phi^3}{6}\right) = \mu\left(\theta - \frac{\theta^3}{6}\right)$. It is clear that, if ϕ is doubled, the difference between the two equations much more than doubles.

The Sine Condition. Even if a lens system is designed, so as to produce an image of a point free from aberration when a wide angle pencil is used, it does not follow that it will give under the same conditions a clear image of an element of area surrounding the point. For this there is necessary an additional condition, called the sine condition, connecting the sines of the angles of divergence of conjugate rays through the point and its image.

Let Qq be a sharp image of a small object Pp, produced by means of a wide angle pencil. Let μ_1 be the index of refraction of the medium

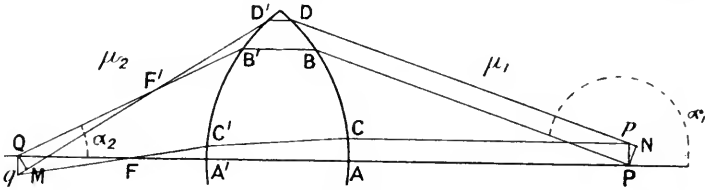


FIG. 63.

in which the object is situated, and μ_2 the index of refraction of the medium in which the image is situated.

Draw through P two rays, one, PA, along the axis and one, PB, making a large angle α_1 with the positive direction of the axis, to meet after refraction in Q. Similarly draw through p two rays, one, pC, parallel to the axis and one, pD, parallel to PB, to meet after refraction in q. The object and image are so small that QB′ and qD′ may be supposed to make the same angle α_2 with QA′. From P draw PN perpendicular to pD and from Q draw QM perpendicular to qD′.

Since F is the focus of the rays pC and PA the path pCC′F is optically equal to the path PAA′F. Since Qq is small and perpendicular to QF, QF = qF to the first order of small quantities. Consequently PAA′FQ = pCC′Fq optically.

Since Q is the image of P, PAA′FQ = PBB′F′Q optically, and since q is the image of p, pCC′Fq = pDD′F′q optically. But PAA′FQ = pCC′Fq optically. Hence PBB′F′Q = pDD′F′q optically.

Since F′ is the focus of the rays pD and PB, the paths NDD′F′ and PBB′F′ are optically equal. Also since QM is small and perpendicular to qD′, F′Q = F′M to the first order of small quantities. Hence NDD′F′M = PBB′F′Q optically.

Combining this result with the former one pDD′F′q = NDD′F′M optically. The parts at the ends, pN and qM, must therefore have the same optical length, i.e.

$$\mu_1 p\text{N} = \mu_2 q\text{M}.$$

Let Pp = y_1 and Qq = y_2. Then pN = Pp sin pPN = $y_1 \sin \alpha_1$ and qM = Qq sin qQM = $y_2 \sin \alpha_2$. Hence

$$\mu_1 y_1 \sin \alpha_1 = \mu_2 y_2 \sin \alpha_2,$$

which is the sine condition. The above method of proof is due to Hockin. When the angles are small, the sine condition coincides with Helmholtz's magnification law. The proof of the latter holds, of course, only for small angles.

A thin pencil diverging from a point on the axis of a lens and falling obliquely on an element of its surface is refracted astigmatically by the lens and forms two focal lines as in the corresponding case of the concave mirror.

Chromatic Aberration. Hitherto we have assumed in dealing with lenses that the light is monochromatic. The index of refraction of all substances varies with the colour or wave-length of the light. It is usual in the tables to specify the index of refraction of a glass by giving its values for the Fraunhofer lines; thus, for example, the following table gives the index of refraction of two of Messrs. Chances' glasses:—

Factory No.	Name.	μ_D.	$\mu_D - \mu_C$.	$\mu_F - \mu_D$.	$\mu_G - \mu_F$.	$\nu = \frac{\mu_F - \mu_C}{\mu_D - 1}$.
605	Hard Crown	1·5175	·00252	·00604	·00484	·01654
361	Dense Flint	1·6214	·00491	·01231	·01046	·02771

The suffixes refer to the Fraunhofer lines, the wave-lengths and colours of those employed being as follows:—

	Colour.	Wave-Length.
C	Red	$6{\cdot}563\ 10^{-5}$ cms.
D	Yellow	5·893 ,,
F	Blue	4·862 ,,
G	Violet	4·308 ,,

The indices of refraction of all glasses increase from the red to the violet.

The focal length of a thin lens is given by

$$\frac{1}{f} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad . \quad . \quad . \quad . \quad (34)$$

If the different values of μ are substituted in this formula, we find that the focal length decreases as we pass along the spectrum from the red to the violet. Thus if a thin convex lens be used to form a real image of a white object, it forms a series of coloured images of slightly different size, situated at different distances from the lens, the violet image being nearest the lens and the red image farthest away. The yellowish-green is the brightest part of the spectrum; consequently if the image be focussed on a screen we unconsciously make the yellowish-green image sharp. Then, superimposed on it, are the other images all somewhat out of focus. The general effect is to give a blurred white image. This blurring of the image, due to the index of refraction being different for different colours, is termed chromatic aberration.

If the screen be moved out of focus towards the lens, the edge of the image becomes tinged with red; if it be moved out of focus the other way, the edge of the image becomes tinged with blue. For in the first case the red image is more out of focus, each point on the edge forming a red disc, and these discs extend beyond the blue point images, and in the second case these conditions are reversed.

Denoting the focal lengths for the lines C and F by f_C and f_F we have

$$\frac{1}{f_C} = (\mu_C - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right),$$

$$\frac{1}{f_F} = (\mu_F - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right),$$

and by subtraction

$$\frac{1}{f_F} - \frac{1}{f_C} = (\mu_F - \mu_C)\left(\frac{1}{r_1} - \frac{1}{r_2}\right).$$

Substitute for $\left(\frac{1}{r_1} - \frac{1}{r_2}\right)$ from equation (34). Then

$$\frac{1}{f_F} - \frac{1}{f_C} = \frac{\mu_F - \mu_C}{(\mu - 1)f}$$

or

$$\frac{f_C - f_F}{f_C f_F} = \frac{\mu_F - \mu_C}{(\mu - 1)f},$$

where μ and f denote any corresponding values of the index of refraction and focal length. Let us now suppose that f is intermediate in value between f_C and f_F so that we may write as an approximation f^2 for $f_C f_F$. Then

$$\frac{f_C - f_F}{f} = \frac{\mu_F - \mu_C}{\mu - 1} \quad . \quad . \quad . \quad . \quad . \quad (35)$$

If the object is at infinity, the expression on the left gives the distance between the red and blue images divided by the focal length of the lens, and can be regarded as a measure of the chromatic aberration. Write

$$\nu = \frac{\mu_F - \mu_C}{\mu - 1};$$

this quantity is known as the dispersive power of the glass for the lines F and C.

In the case of a convex lens the red image is always farthest from the side of the incident light; in the case of a concave lens, it is always nearest the side of the incident light. The question thus arises as to whether it is possible to combine a convex and a concave lens, so as to make the chromatic aberration in the one neutralize the same defect in the other. To this question Newton answered no. He believed that $\mu_F - \mu_C$ for a glass was always proportional to its refractivity, i.e. to $\mu - 1$, and that consequently a system free from aberration would act simply like a piece of glass with plane parallel sides, or in other words, that the blurring of the different colours was essential to the image-forming properties of the lens.

This view was, of course, wrong, but he was led to adopt it by his own observations on crown glass and water where the proportionality holds approximately.

Let us suppose that two thin coaxal lenses are placed in contact, that the focal length and dispersive power of the first are denoted by f and ν, while the focal length and dispersive power of the second are denoted by f' and ν'. Then F, the focal length of the combination, is given for the C line by

$$\frac{1}{F_C} = \frac{1}{f_C} + \frac{1}{f'_C}$$

and for the F line by

$$\frac{1}{F_F} = \frac{1}{f_F} + \frac{1}{f'_F}.$$

By subtraction we obtain

$$\frac{1}{F_F} - \frac{1}{F_C} = \frac{1}{f_F} - \frac{1}{f_C} + \frac{1}{f'_F} - \frac{1}{f'_C}$$

or

$$\frac{F_C - F_F}{F^2} = \frac{f_C - f_F}{f^2} + \frac{f'_C - f'_F}{f'^2},$$

writing F^2 for $F_C F_F$, f^2 for $f_C f_F$ and f'^2 for $f'_C f'_F$ in the denominators. By means of (35) this last equation may be written

$$\frac{F_C - F_F}{F^2} = \frac{\nu}{f} + \frac{\nu'}{f'}.$$

Thus if for two colours the relation

$$\frac{\nu}{f} + \frac{\nu'}{f'} = 0 \quad . \quad . \quad . \quad . \quad . \quad (36)$$

holds, the images formed by light of these colours coincide in size and position, and the combination is said to be achromatic for these two colours.

As an example on equation (36) let us suppose that a convex lens of focal length 35 cms., achromatic for the two lines C and F, is to be made from the two glasses, the data for which is given on p. 61. Then, if the light is coming from the right, we have the equations

$$-\frac{1}{35} = \frac{1}{f} + \frac{1}{f'},$$

$$0 = \frac{\cdot 01654}{f} + \frac{\cdot 02771}{f'},$$

the solution of which is $f = -14{\cdot}10$ cms., $f' = 23{\cdot}62$ cms. The combination must therefore consist of a convex lens of the crown glass of focal length 14·10 cms. and a concave lens of the flint glass of focal length 23·62 cms. The combination is, of course, achromatic only for the region for which it is calculated.

If we assume that the above values of f and f' are for the D line and calculate the values for the other colours we obtain the results in the following table :—

	f.	f'.	F.
C	14·17 cms.	23·80 cms.	35·02 cms.
D	14·10 ,,	23·62 ,,	34·99 ,,
F	13·94 ,,	23·15 ,,	35·05 ,,
G	13·81 ,,	22·79 ,,	35·09 ,,

It will be observed that F varies much less with the wave-length than f or f'. If the combination is for use in photography with ordinary plates, it should be corrected for the violet and ultra-violet.

In the above achromatic doublet only the focal lengths, not the radii of curvature of the two component lenses are prescribed. We have consequently a radius of curvature in each lens at our disposal. According to p. 59 the spherical aberration depends on the ratio of the radii of curvature of the two outside faces; we can choose this ratio so as to make the spherical aberration a minimum. The remaining relation between the radii of curvature can be used to make the inside curvatures of the two lenses the same. They can then be cemented together, and an air space with its consequent loss of light by reflection avoided.

If the convex component of the above doublet is equi-convex its radius of curvature is 14·6 cms., and if the concave component is plano-concave its radius of curvature is also about 14·6 cms. The two lenses thus fit together and form a plano-convex combination. The latter shape, as has been seen, has a very low spherical aberration for parallel light. The object glasses of small telescopes are therefore very often made of an equi-convex crown glass lens cemented to a plano-concave flint glass lens, mounted so that the light is incident first on the crown glass lens.

Let us suppose that the diameter of the above achromatic doublet is 3·5 cms. Then its spherical aberration can be calculated by multiplying the values on p. 59 by $\left(\frac{3\cdot5}{2}\right)^2\Big/35$, i.e. by ·087. For an index of refraction 1·5 this gives 1·0 mm., and for an index of refraction 2·0 it gives ·4 mm.

If three thin lenses made of glasses of different dispersive powers be combined into a single lens, the images formed by three different colours may be made to coincide.

For, proceeding in the same way as on p. 63, we can obtain the equation

$$\frac{F_C - F_F}{F^2} = \frac{\nu}{f} + \frac{\nu'}{f'} + \frac{\nu''}{f''},$$

where f'' is the focal length of the additional lens and ν, ν', ν'' are the dispersive powers for the two lines C and F. Similarly, if ω, ω', ω'' denote the dispersive powers for the lines F and G, we have

$$\frac{F_F - F_G}{F^2} = \frac{\omega}{f} + \frac{\omega'}{f'} + \frac{\omega''}{f''},$$

and if f, f', f'' be chosen to satisfy the equations

$$\left.\begin{aligned} \frac{1}{\mathsf{F}} &= \frac{1}{f} + \frac{1}{f'} + \frac{1}{f''} \\ 0 &= \frac{\nu}{f} + \frac{\nu'}{f'} + \frac{\nu''}{f''} \\ 0 &= \frac{\omega}{f} + \frac{\omega'}{f'} + \frac{\omega''}{f''} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (37)$$

it is clear that the lens formed by the three thin lenses is achromatic for the lines C, F, and G. Of course, owing to their form, equations (37) always give real values for f, f', and f''.

The chromatic error remaining in the image after it has been achromatised for two colours is often referred to as "secondary spectrum". It can be diminished considerably by using some of the new glasses made in Jena. They appear, however, to offer difficulties in manufacture and to be not very durable.

Chromatic Aberration of Two Thin Lenses separated by a Finite Interval. Let f, f' be the focal lengths of the two lenses and a the distance between them. Then by p. 46 F, the equivalent focal length of the combination, is given by

$$\frac{1}{\mathsf{F}} = \frac{f + f' + a}{ff'}.$$

To make the system achromatic for two colours for any position of the object, it would be necessary not only to give F the same value but also to give the equivalent planes the same positions for the two colours. This is impossible. It is even impossible to make the system achromatic for one position of the object.

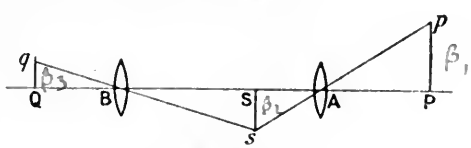


FIG. 64.

For, let $\mathsf{Q}q$ be an image formed of $\mathsf{P}p$ by the two lenses A and B, $\mathsf{S}s$ being the intermediate image formed by the lens A. Let $\mathsf{P}p = \beta_1$, $\mathsf{S}s = \beta_2$, and $\mathsf{Q}q = \beta_3$. Let $\mathsf{AP} = u$, $\mathsf{AS} = v$, $\mathsf{BS} = u'$, and $\mathsf{BQ} = v'$. Then

$$\frac{\beta_1}{\beta_2} = \frac{u}{v}, \quad \frac{\beta_2}{\beta_3} = \frac{u'}{v'},$$

and therefore

$$\frac{\beta_3}{\beta_1} = \frac{vv'}{uu'}.$$

If the images formed by the two colours coincide in position and size, u, v' and β_3/β_1 are the same for both colours. Hence v/u' is the same for both colours. But the distance BA is fixed. Therefore the point S is the same for both colours, or in other words the lenses A and B must be achromatic themselves.

Let the equivalent focal lengths of the combination for the C and F lines be denoted by F_C and F_F. Then

$$\frac{1}{F_C} = \frac{1}{f_C} + \frac{1}{f'_C} + \frac{a}{f_C f'_C}$$

and

$$\frac{1}{F_F} = \frac{1}{f_F} + \frac{1}{f'_F} + \frac{a}{f_F f'_F},$$

whence by subtraction

$$\frac{1}{F_F} - \frac{1}{F_C} = \frac{1}{f_F} - \frac{1}{f_C} + \frac{1}{f'_F} - \frac{1}{f'_C} + \frac{a(f_C f'_C - f_F f'_F)}{f_F f'_F f_C f'_C}.$$

The last term is equal to

$$\frac{a(f_C f'_C - f_C f'_F + f_C f'_F - f_F f'_F)}{f_F f'_F f_C f'_C} = \frac{a(f_C\{f'_C - f'_F\} + f'_F\{f_C - f_F\})}{f_F f'_F f_C f'_C}$$

$$= \frac{a}{f_F}\left(\frac{1}{f'_F} - \frac{1}{f'_C}\right) + \frac{a}{f'_C}\left(\frac{1}{f_F} - \frac{1}{f_C}\right).$$

Hence in the same way as on p. 63

$$\frac{F_C - F_F}{F^2} = \frac{\nu}{f} + \frac{\nu'}{f'} + \frac{a(\nu + \nu')}{ff'}.$$

Suppose now that the two lenses are made of the same glass and that $\nu' = \nu$. Then

$$\frac{F_C - F_F}{F^2} = \frac{\nu}{ff'}(f + f' + 2a),$$

i.e. if $f + f' + 2a = 0$, the equivalent focal length of the system is constant, no matter what the value of ν is.

Thus if a system of two thin coaxal lenses made of the same glass is mounted with the distance between them numerically equal to half the sum of their focal lengths, the focal length of the combination is the same, not only for any two, but for all colours.

EXAMPLES.

(1) A parallel beam is incident on a concave spherical mirror. Draw the caustic curve.

(2) A concave spherical mirror has a radius of curvature of 50 cms. A point object is situated on the axis 90 cms. from the surface of the mirror. Calculate the aberration for the rays which meet the surface of the mirror 2, 4, 6 and 8 cms. from the axis.

(3) Investigate the case of astigmatic reflection at a convex mirror in exactly the same way as astigmatic reflection at a concave mirror is treated in the present chapter.

(4) A concave mirror has a radius of curvature of 40 cms. and the diameter of its rim is 5 cms. A point object is 50 cms. distant from it in a direction making an angle of 45° with the axis of the mirror. Find the positions of the focal lines and calculate their lengths, i.e. the lengths of the lines TT′ and SQ in fig. 60.

(5) A small pencil of light diverges from a point in a medium of refractive index μ, is incident obliquely on the plane surface of this medium and then passes into air. Show that it appears to diverge from two focal lines in the first medium distant v_1 and v_2 from the point of refraction, where v_1 and v_2 are given by

$$v_2 = \frac{u}{\mu}, \quad v_1 = \frac{u \cos^2 r}{\mu \cos^2 i};$$

u is the distance of the point object from the point of refraction and i and r are the angles of incidence and refraction.

(The two focal lines are virtual in this case and hence cannot be received on a screen. But if a convex lens is used to form a real image of them, their existence can be demonstrated very neatly. The best way of producing them is to employ a block of plate glass 4 inches by 3 inches by $\frac{3}{4}$ inch or thereabouts, such as is used in elementary experiments on refraction, and as point source to take a small hole in a metal plate placed close up to one of the ends of the block with the filament of a glow lamp directly behind the hole. The plate is arranged so that the path of the rays in it is as long as possible. After emerging from the plate the rays are received by a convex lens which focusses them on a screen. The image is a straight line. By displacing the screen another straight line comes into focus at right angles to the first.)

(6) Show that equation (28) reduces to (24) on the substitution of -1 for μ. Can all formulæ dealing with reflection at a spherical surface be treated in this way as a particular case of refraction at the same surface?

(7) A convex lens of focal length 50 cms., achromatic for the lines D and F, is to be made from the two glasses, the data for which are given on p. 61. Find the focal lengths of the components and calculate how far the combination is out for the wave-lengths C and G.

(8) Light diverging from a point is refracted at a plane surface; prove that the caustic curve is the evolute of a hyperbola if the point is in the less dense medium, and the evolute of an ellipse if the point is in the more dense medium.

(9) The radii of curvature of both faces of a thin convex lens are the same and it is made of glass of refractive index 1·52. Derive an expression for the spherical aberration when the object is at a distance of twice the focal length from the lens, and verify it by experiment.

(10) A table of the refractive indices of a number of optical glasses is given at the end of the book. Which two make the best achromatic doublet for the region of the spectrum from C to G? Other things being equal, these glasses are to be avoided which involve small values of f and consequently steep curves. Illustrate your result by numbers, assuming that the focal length of the doublet is to be 100 cms.

CHAPTER V.

ON DETERMINING THE CONSTANTS OF MIRRORS AND LENSES.

Optical Bench. The apparatus used most in physical laboratories for determining the focal lengths of lenses is the optical bench. Optical benches can be divided into two classes, those with wooden bases and fittings and those with metal bases and fittings. The latter are much more expensive, are usually too elaborate, accuracy and labour being wasted on their construction where it is not wanted, and although indispensable for certain special uses, are not to be recommended for general purposes.

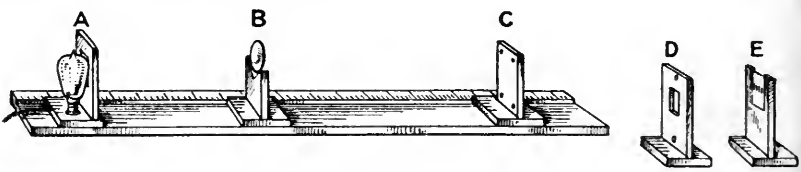


FIG. 65.

Fig. 65 shows a very useful and simple wooden bench together with its fittings. The base is mahogany. At the side is a scale two metres long for reading the positions of the pieces. The stand A carries an incandescent electric lamp, in front of which is a wooden upright with a rectangular hole in it, across which are stretched cross-wires. These cross-wires are the object, and it is usually advisable to fasten a piece of tissue paper with drawing pins between the hole and the lamp to give a more uniform background to the object. Stand B is for carrying the lens or mirror; it has a V-shaped top with a groove in it. C is a screen for receiving the image, consisting of an upright with a piece of paper fastened on with drawing pins, D a similar screen with a hole in it also used for receiving images, and E a stand carrying a square of ground glass with a scale on it. The positions of the various stands can be read by marks on their bases. These marks may not be placed accurately, and so before taking a series of readings of the distance between cross-wires and lens, for example, it is usual to place a rod of known length with one end touching the cross-wires and the other touching the lens and to compare the distance as read by the scale with the known length of the rod. If there is any difference, it should be employed as a correction to each reading of the series.

Focal Length of a Convex Lens. To find the focal length of a convex lens it is simply put on the stand B and an image of the cross-wires focussed on the screen. Then u and v, the distances of the cross-wires and screen from the lens, are measured and f is calculated by the formula

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}.$$

The length of one of the cross-wires and the length of its image can at the same time be measured by callipers and thus the linear magnification found. It should, of course, be equal to v/u.

Another way of finding the focal length is to place the lens at the middle of the bench and arrange the cross-wires and image screen at equal distances from it. Then, in general, there will be no image formed. If now the distances of the cross-wires and image screen from the lens be gradually increased or decreased, their values being always kept equal to one another, positions will eventually be reached in which a sharp image is formed on the screen. Then u is numerically equal to v and each is numerically equal to $2f$. If d be the distance between the cross-wires and image screen, f is numerically equal to $d/4$.

If d is numerically less than $4f$, no real image is formed. If d is numerically greater than $4f$, for every given position of the cross-wires and image screen two positions can be found for the lens in which it gives a real image. This follows simply from the nature of the formulæ, for they can be written arithmetically

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{f}, \quad u + v = d,$$

and thus u and v can be interchanged without their form being altered. Thus, if originally $u = 10$ cms. and $v = 15$ cms., and if the cross-wires be kept fixed and the lens be moved out another 5 cms. so that u becomes 15 cms., v becomes 10 cms. and an image again appears on the screen. The magnification in the one case is the reciprocal of the magnification in the other.

If a be the distance between the two positions of the lens, $a = v - u$. But $d = v + u$; hence $2v = d + a$, $2u = d - a$, and

$$\frac{1}{f} = \frac{1}{v} + \frac{1}{u} = \frac{2}{d+a} + \frac{2}{d-a} = \frac{4d}{d^2 - a^2}$$

or

$$f = \frac{d^2 - a^2}{4d}.$$

Hence, if d and a be measured, f can be calculated. This method is called the double-position method.

A fourth method is simply to measure the distance from the lens of the image of a lamp situated at a distance, which is great in comparison with the focal length of the lens.

Focal Length of a Concave Lens. If the object is real, the image produced by a concave lens is always virtual; if the image is real, the object is always virtual. Consequently the focal length of a concave lens can never be determined by using it alone on the optical bench; an auxiliary convex lens must be employed.

There are two simple methods for determining the focal length of a concave lens. It is necessary for the first of these that the auxiliary convex lens should be more powerful, i.e. have a shorter focal length than the concave lens. For the second method any convex lens will do.

In the first method the concave lens and auxiliary convex lens are placed together in close contact on stand B. The combination acts as a convex lens and its focal length F is determined. The focal length, f, of the convex lens alone is next determined. Then f', the focal length of the concave lens, is given by the algebraic formula

$$\frac{1}{F} = \frac{1}{f} + \frac{1}{f'}.$$

The sign of f' comes out different from the signs of F and f.

In the second method the convex lens is first used alone to form an image Q of the cross-wires on the screen. The concave lens is

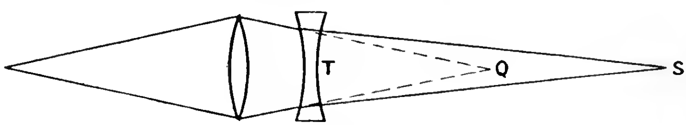


Fig. 66.

next mounted on another stand similar to B and inserted at T between the convex lens and screen. The rays after passing through it become less convergent, and the screen has to be moved to S to bring the image again into focus. Thus the focal length can be determined by the algebraic formula

$$\frac{1}{f'} = \frac{1}{v} - \frac{1}{u},$$

where TS $= v$ and TQ $= u$.

In this case the object is virtual and the image real.

Focal Length of a Concave Mirror. The focal length of a concave mirror is given by the formula

$$\frac{1}{f} = \frac{2}{r} = \frac{1}{u} + \frac{1}{v}.$$

Hence f may be found simply by measuring r with the spherometer and dividing the result by 2.

It may be determined on the optical bench by using the screen D with the hole in it. This screen is placed between the mirror and cross-wires. The rays from the cross-wires fall on the mirror through the hole on the screen and are reflected to form an image on the screen.

The mirror is given a slight tilt, otherwise the image would fall on the hole itself. Then, if u and v are measured, f can be calculated.

If the lamp is taken to a very great distance, v, of course, becomes equal to f.

Another method consists in sticking a pin in the rectangular opening in stand D, as nearly in the plane of the paper as possible, and reversing the stand A so as to throw more light on the pin. Then, if the image of the pin be formed on the screen, $u = v = r$ and f is half the distance of the screen from the mirror.

Focal Length of a Mirror. The focal length or radius of curvature of a concave spherical mirror can also be determined by the following method, which is of interest because, unlike the methods in the preceding section, it can also be applied to a convex spherical mirror.

A paper scale about 50 cms. long pasted on wood is mounted horizontally directly in front of the concave mirror (fig. 67) about

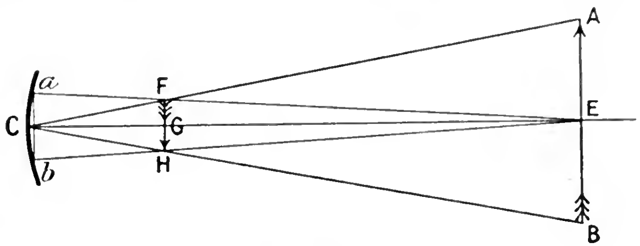


Fig. 67.

the same height above the table and two or three metres from it. Two metal strips A and B slide along the front of this scale. An image of AB, i.e. the distance between the strips, is produced at G, and if FH be this image, the points C, F, and A and the points C, H, and B are collinear, since the lengths of the object and image are in the ratio of their distances from the mirror. A telescope is placed with its object glass at E, the centre of AB, and focussed on FH. A small scale is fixed in front of and in contact with the mirror which is tilted so that, when viewed through the telescope, the image FH appears parallel to and touching the edge of this scale. The distance between the metal strips is adjusted so that the image FH appears to coincide with an exact number of divisions on this scale, l cms., say. Let r be the radius of curvature of the mirror, let $AB = L$ and let $CE = D$; then

$$r = \frac{2lD}{L + 2l}.$$

If the mirror is convex

$$r = \frac{2lD}{L - 2l}.$$

This is the easiest of all the optical methods for determining the radius of curvature of a convex mirror. It should be noted that the small scale and the image must be in focus at the same time ; hence D must be large in comparison with r.

To prove the formula for the case of the concave mirror let CG $= v$; l, of course, is equal to ab. Then

$$\frac{L}{l} = \frac{L}{FH} \cdot \frac{FH}{l} = \frac{D}{v} \cdot \frac{D - v}{D} = \frac{D}{v} - 1 \quad . \quad . \quad . \quad . \quad (38)$$

by similar triangles. But

$$\frac{1}{D} + \frac{1}{v} = \frac{2}{r};$$

therefore

$$1 + \frac{D}{v} = \frac{2D}{r}. \quad . \quad . \quad . \quad . \quad . \quad (39)$$

Hence eliminating D/v between equations (38) and (39) we obtain

$$2 + \frac{L}{l} = \frac{2D}{r}$$

or

$$r = \frac{2lD}{L + 2l},$$

the required relation.

Cylindrical Lenses. Let us suppose that the surfaces of a thin piece of glass are cylinders, the axes of which are parallel. Such a piece of glass is called a cylindrical lens.

If a section be made of it by a plane at right angles to both axes, it has the same shape as the section of a lens with spherical surfaces. Consequently any pencil of rays diverging from a point in this plane is brought to a focus by the lens. On the other hand, the section by any plane parallel to the axes is the same as the section of a thin plate with parallel sides. Consequently a pencil of rays in this plane is unaffected by the lens.

A cylindrical lens thus forms images only of lines parallel to the axes of its surfaces. If such a lens is placed on the optical bench with its axes parallel to one of the cross-wires it forms a sharp image of that wire but no image at all of the other wire. As far as the other wire is concerned the lens acts merely as a plane parallel plate.

With this limitation the focal lengths of cylindrical lenses and mirrors can be determined on the optical bench in the same way as the focal lengths of spherical lenses and mirrors.

Magnification Methods for Determining Focal Lengths. The methods of determining the focal lengths of convex lenses given hitherto are suitable only for thin lenses. The three methods which follow are suitable also for thick lenses or systems of lenses :—

(1) Form an image on the optical bench. Denote the magnification by m. Keep the object and image screen fixed and displace the lens through a distance d, so as to give a clear image again. Let the magnification now be m'. Then

$$f = \frac{d}{m - m'}.$$

(2) Abbe's Method. Make the first magnification, m, unity or less, keep the lens fixed, displace the image screen a distance d away from the lens and move the object until focus is again obtained. Let the magnification in the second case be m'. Then

$$f = \frac{d}{m - m'}.$$

(3) Same as (2), only d in this case is the distance moved by the object and

$$f = \frac{d}{\frac{1}{m} - \frac{1}{m'}}.$$

To carry out the above methods on the bench shown on p. 68, two lantern slides were made by photography from a white paper scale, but in mounting them pieces of ground glass were used instead of the cover glass, the ground side being next the photographic film. The two scales thus prepared were, of course, identical. One of them, mounted on the stand E, was used as object. It was illuminated by placing lamp A behind it, the lamp being turned round so that the wooden upright did not come between. The other scale mounted on a similar stand was used as image screen. The magnification was found by getting the image of the one scale directly below the other and finding how many divisions on the one corresponded to the whole length of the other.

One advantage of the above methods is that the only length involved, d, is in every case not a distance between two objects but a displacement of one object, and the latter can be measured with much greater accuracy.

To prove the three formulæ let u and v in each case denote the distances of the object and image measured from their respective principal planes before the displacement, and let u' and v' denote the same distances after the displacement.

Then in each case we have

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}, \quad \frac{1}{v'} - \frac{1}{u'} = \frac{1}{f},$$
$$m = v/u, \; m' = v'/u'.$$

In the first and second cases we have

$$d = v' - v = f(1 - m') - f(1 - m) = f(m - m'),$$

i.e.

$$f = \frac{d}{m - m'}.$$

In the third case

$$d = u - u' = f\left(\frac{1}{m} - 1\right) - f\left(\frac{1}{m'} - 1\right) = f\left(\frac{1}{m} - \frac{1}{m'}\right),$$

i.e.

$$f = \frac{d}{\frac{1}{m} - \frac{1}{m'}}.$$

The Nodal Slide. The focal planes of a system of lenses can be found very easily by simply letting a beam of parallel light fall on it from each side in succession and noting where the image is formed. Then, when the focal length is determined, the principal planes and nodal points can be found simply by measuring off their distances from the focal planes. There is, however, an elegant method of determining the nodal points directly.

The property of the nodal points, it will be remembered, was, that if a ray of light passed through the one, its conjugate passed through the other, and further was always parallel to the incident ray. Suppose now a beam of parallel rays is incident on the system in the object space, and that the system is pivoted so that it can rotate about the vertical through the nodal point of the image space. As it does so, the nodal point of the object space describes a short arc and so different rays in turn pass through it. But the incident rays are all parallel, so the direction of the ray through the nodal point of the object space is always the same. Consequently the direction of the ray through the nodal point of the image space is always the same, and if the image is received on a screen, it remains stationary when the system is rotated. The distance from the screen to the axis of rotation gives, of course, the focal length of the system.

To put the method into practice the lens system is mounted on a stand which is placed on a turntable, along one side of which is fixed a scale for reading the position of the stand. This apparatus is known as a nodal slide. Then the turntable is rotated for different positions of the stand on the scale. Unless the axis of rotation passes through the nodal point of the image space, the image moves on the screen. When the right position is passed, the direction of the motion of the image on the screen changes.

More Accurate Method of Determining Focal Length. Focal lengths are usually determined in a Physical Laboratory not for the numerical result, but for the purpose of teaching students the principles of optics. The lenses measured are usually spectacle lenses with a focal length of 20 or 30 cms., because such lenses suit a two-metre optical bench well. The methods for a convex lens on p. 69 give single results which agree to a millimetre for such lenses, but the magnification methods as used on p. 72 are not so accurate. If greater accuracy is really desired or if a method is wanted that will also give, for example, the equivalent focal length of a telescope eyepiece, recourse may be had to a more elaborate apparatus such as the Beck Lens Testing Bench. It ought to be pointed out, however, that by means of a vernier microscope such as is to be found in every laboratory, values of the focal length can be got which will compare favourably with those obtained by any other method.

The method consists in the measurement of the aerial image of an object at a great distance. For camera lenses I have taken a horizontal 50 cm. scale distant 12 metres. It was illuminated by a lamp or even by Na light. For microscope objectives the distance should be much less. The vernier microscope should have both a horizontal and a vertical motion. The lens is mounted on a stand in front of it and the length of the image measured. If m is the magnification and d the distance of the object from the image, and if u and v are used arithmetically,

$$m = \frac{v}{u}, \quad d = u + v = u(1 + m)$$

and

$$f = \frac{uv}{u + v} = \frac{mu}{1 + m} = \frac{md}{(1 + m)^2},$$

m being usually so small that the denominator may be made unity.

A beauty of this method is that the percentage error of measurement is about the same for the lengths of the object and the image and for the distance d.

So far the distance between the principal planes has been neglected, but from the first value of f it can be estimated,* the result subtracted from d, and a more accurate calculation made. d is so great that it does not require to be known very accurately.

The method can be made applicable to a long focus lens by using distant objects and taking the angle between them with a sextant. Then, if the image length is l and the angle 2α, f is simply $l/(2 \tan \alpha)$.

Investigation of the Aberrations of a Lens. In deriving the formula connecting the positions of the object and image formed by a lens system, the angle which each ray makes with the axis is assumed to be so small, that it may be put equal to its sine and that its cosine may be put equal to unity. If the second terms in the expansions for the sine and cosine be taken into consideration, the passage of the rays through the system may still be treated from a general standpoint, and it is found that when the object is a point, the rays in the image space no longer converge to a point but to a small region. If the angles made with the axis are so great that two terms in the expansion are not a sufficient approximation, there is no elegant method of treatment; the paths of the rays in the image space must be found by laborious trigonometrical calculation.

If the rays do not make small angles with the axis, the image is blurred or has defects. In certain particular cases these defects take characteristic forms—spherical aberration, coma, astigmatism, curvature of the surface, distortion—which depend each on a separate constant in the mathematical theory. Thus in the general case these five defects may be regarded as occurring independently and to a varying extent. They are characteristic of the system for a definite position of the object. The system can be corrected for any one of

* I.e. by measuring off a distance f from the aerial image and making a scratch on the lens mounting, then reversing the lens and repeating the operation. The two scratches give the principal planes since d is so very great.

them by making the appropriate constant vanish, and the image-forming properties of the system can be investigated by testing for each of them separately. It should be stated that there are at present no generally recognised methods of testing or of measuring these defects numerically.

Of course in addition to the above there may be defects due to the surfaces of the lenses not being true or to the separate components being badly centred. There is also chromatic aberration.

As excellence in one respect is gained often at the expense of comparative failure in another respect, lenses should, of course, be tested only under the conditions for which they are to be used and for the defects which are important under these conditions.

For measuring the aberrations of lenses properly an apparatus such as the Beck Testing Bench is absolutely necessary. However two simple arrangements will be described here which are of some use when the aberrations are large.

For the first of these there is necessary a vernier microscope which can be racked forward in the direction of its length. Vertical motion and horizontal motion at right angles to the length are a great convenience, but not absolutely necessary. As source of light a circular hole of 1 mm. diameter drilled in a thin sheet of aluminium is used. It is held in front of a lamp and placed at a distance of about 12 metres from the microscope. The lens to be tested forms an aerial image of this hole and this image is examined with the microscope.

To measure the chromatic aberration a piece of red glass is placed behind the hole and the image focussed. Then the red glass is replaced by a piece of green and a piece of blue, which together give an approximately monochromatic green, and the microscope racked forward to focus the image a second time. The distance between the two images, l, divided by the focal length is an approximate measure of the chromatic aberration.

To measure the spherical aberration a screen with two holes in it is used, one to transmit the rays that pass through the centre of the lens and the other to transmit the rays that pass through a marginal zone. These stops are placed in succession in front of the lens. If l is the distance between the images in the two cases and θ half the angle subtended at the image by the mean diameter of the marginal zone, the expression

$$\frac{l}{f\theta^2}$$

may be taken as a measure of the spherical aberration, since for an uncorrected lens the distance between the images is proportional to θ^2.

To measure the astigmatism the lens is placed so that its axis forms an angle θ of about 10° with the straight line from the microscope to the source. Then, instead of a point image, two line images may be seen at right angles to one another and distant l apart along the axis of the microscope. They may however be obscured by the other defects. The distance between these lines should be proportional to θ^2 for small values of θ. Hence the astigmatism may be approximately measured by

$$\frac{l}{f\theta^2}.$$

If a lens forms an image of a point source not on its axis, the central zone of the lens gives a point image and the other zones may give ring images of increasing diameter according to the distance of the zone from the centre of the lens, even if the lens is corrected for spherical aberration. These rings are not concentric but situated as in fig. 68, the radius of each equalling half the distance of its centre from the point image. The result is to give an egg-shaped patch of light with the narrow end much brighter than the other. This defect is called coma and vanishes if the system obeys the sine condition.

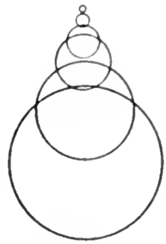


FIG. 68.

The curvature and distortion of the image formed by a spectacle lens can be measured very well by the apparatus shown in fig. 69, which consists of an optical bench with a cross-piece at one end. A stand carrying cross-wires and illuminated by a lamp can move along the cross-piece. Instead of the usual screen a long white scale is used for receiving the image, and

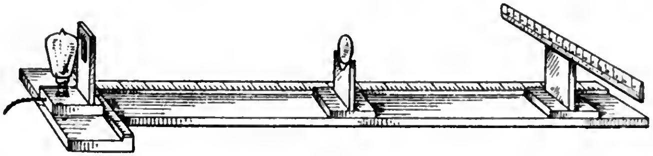


FIG. 69.

this scale and the stand carrying the lens move along the bench itself. The bench and cross-piece are both provided with millimetre scales.

The lamp is first placed so that the line joining object and image is exactly parallel to the axis of the bench. Then if the lamp is moved a distance a along the cross-piece, the image is displaced a distance b along the scale, and at the same time the latter has to be moved a distance c along the bench to restore focus. The distortion is measured by the deviation from proportionality of b to a and the curvature is proportional to c/b^2.

EXAMPLES.

(1) An object 2 cms. high is on the axis of a thin convex lens of focal length 20 cms. at a distance of 50 cms. to the left of it. A thin concave lens of focal length 45 cms. is placed 10 cms. to the right of and on the same axis as the convex lens. Find the position and magnitude of the final image.

(2) If on the optical bench a thin lens is set not quite at right angles to the axis of the bench but rotated through a small angle about a vertical axis, then the cross-wires used as object, which are supposed vertical and hori-

zontal, do not come into focus at the same time. If the vertical wire is in focus, the image screen has to be displaced through a small distance in order to bring the horizontal wire into focus. Determine experimentally how this displacement varies with the angle between the axes of the lens and bench. The angle is best measured with protractors.

(3) A student is given a convex and a concave lens and required to determine their focal lengths. Their rims have equal diameters, but a piece of the concave lens, perhaps a sixth of its whole area, has been accidentally broken off. The focal length of the convex lens is easily found to be 24·0 cms., and, as the convex lens appears more powerful than the concave one, the first method of determining the focal length of the latter is employed. But to his astonishment the student finds that the combination always gives two images, a faint one near the lenses and a brighter one farther out. If the faint image is used, it gives a focal length of about 24·3 cms. for the combination, and, if the bright one is taken, the result is about 38·2 cms. Which image should he use and why?

(4) The focal length of a concave mirror is obtained in the usual way by measuring u and v and employing an image screen with an aperture in it, through which the rays fall on the mirror. The mirror has to be tilted, otherwise the image would fall upon the aperture itself. Discuss with the aid of the theory of astigmatic reflection at a concave mirror the magnitude of any error thus introduced, and give numbers for an actual case.

(5) The focal length of a concave mirror is obtained by forming the image of a lamp several metres distant. The result is 32·1 cms. How far away must the lamp be for the result to be accurate to 2 mm.?

(6) A convex spherical lens of focal length 22 cms. is placed in contact with a cylindrical lens, and the combination used to form an image of illuminated cross-wires, which are supposed vertical and horizontal. One face of the cylindrical lens is plane and the axis of the cylindrical surface is vertical. Find where the images of the separate cross-wires are formed, given that the object distance is 50 cms., and that the cylindrical lens is (*a*) convex and of focal length 40 cms., (*b*) concave and of focal length 40 cms.

(7) The formulæ given on p. 71 for the radius of curvature of a spherical mirror become much simpler if l can be neglected in comparison with L. Show that this assumption is equivalent to supposing that the small scale is in the same plane as the image, and derive the simplified formulæ from first principles on this supposition.

(8) Two cylindrical lenses, which have each one plane face, are placed in contact and used to form a real image of illuminated cross-wires distant 30 cms. The axes of the cylindrical surfaces are at right angles to one another and each parallel to one of the cross-wires, and images of a cross-wire are formed at distances of 40 and 60 cms. from the lens combination. What are the focal lengths of the lenses, and where would the image be formed if the axes of both cylindrical surfaces were parallel to the same cross-wire?

(9) A convex lens is mounted on an optical bench and used to form an image on a screen. The magnification is 2·41. The object and image screen are then both kept fixed and the lens displaced a distance 10 cms. towards the image when a sharp image again appears, this time of magnification ·415. Calculate the focal length of the lens.

(10) Make n independent determinations (say 8 or 10) of the focal length of a thin convex lens by the first method described in this chapter. Take the mean of the result, take the differences of the individual determinations from the mean, square them, then add their squares; divide the sum by $n(n-1)$ and take the square root of the quotient. The square root is called

the "mean error" of the final result. Multiplication of the mean error by $\frac{2}{3}$ (more properly 0·674) gives the "probable error" of the final result. Multiplication of the mean error and probable error of the result by the square root of the number of observations gives the mean error and probable error of the individual determinations.

For example, the first column of the following table gives 10 determinations of the focal length of a lens, made on the same optical bench by the method referred to :—

f.	δ.	δ^2.
19·91 cms.	− ·244	·05954
20·07 ,,	− ·084	·00706
20·11 ,,	− ·044	·00194
20·16 ,,	+ ·006	·00004
20·14 ,,	− ·014	·00020
20·27 ,,	+ ·116	·01346
20·21 ,,	+ ·056	·00314
20·14 ,,	− ·014	·00020
20·22 ,,	+ ·066	·00436
20·31 ,,	+ ·156	·02434
Mean 20·154		Sum ·11428

The mean error of the result is $\sqrt{\cdot 11428/90} = \cdot 0356$.

The probable error of the result is $\frac{2}{3} \times \cdot 0356 = \cdot 0237$.

The mean error and probable error of a single reading are respectively $\cdot 0356\sqrt{10}$ and $\cdot 0237\sqrt{10}$, i.e. ·113 and ·0749.

The method of obtaining the above results depends on the theory of probability, and they are valid only if the individual readings are distributed about the correct value according to the law of chance. They do not take into account any one-sided errors such as assuming, for example, that the lens employed is infinitely thin and that its principal points coincide.

The probable error of the single determination gives a quantity, above and below which theoretically there should lie an equal number of values of δ. In the above case we have four above, namely, ·244, ·084, ·116, and ·156, and six below, ·044, ·006, ·014, ·056, ·014, and ·066.

CHAPTER VI.

OPTICAL INSTRUMENTS.

Magnifying Glass. The normal eye focusses best on an object distant 10 or 12 inches from it. This distance is called the distance of distinct vision. If we attempt to increase the detail visible by bringing the object nearer, an exertion is required to see it distinctly. Hence the distance of distinct vision is the most favourable position for examining the detail of an object.

Let AB be a convex lens of small focal length placed before the

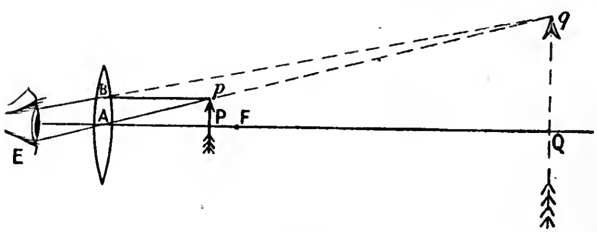


FIG. 70.

eye at E, let F be its focus, and let Pp be an object within its focal length. Then on applying the usual graphical construction, the image is found to be at Qq. All rays proceeding from Pp to the eye, after refraction by the lens, appear to come from Qq.

We have

$$\frac{y_2}{y_1} = \frac{v}{u} = 1 - \frac{v}{f},$$

where y_2 is the length of the image and y_1 is the length of the object. Let us suppose that the image is at the distance of distinct vision. Denote the value of the latter by D. Then since the eye is close to the lens, we can write $v = \text{D}$. Consequently $y_2 = y_1\left(1 - \frac{\text{D}}{f}\right)$ or approximately $-y_1\text{D}/f$, since f is small in comparison with D.

The apparent size of an object depends on the angle it subtends at the eye, but if the object Pp were to be examined by the eye direct, it would have to be placed at Q. The apparent sizes without and with the lens are therefore as the linear dimensions of the object and image, if the latter are small. Hence the magnification produced by the lens is D/f, if the sign is neglected.

A single convex lens used as above is sometimes called a simple microscope.

Instead of a single lens, a combination of thin lenses may be used, for example, two concave meniscus lenses of flint glass separated by a double convex lens of crown glass. Such a combination can be corrected for chromatic aberration, astigmatism, and distortion.

The Astronomical Telescope. The optical system of an astronomical telescope consists of two parts, one called the object glass, which produces a real image of a distant object, and another, the eyepiece, which produces in turn an enlarged virtual image of this image.

It is illustrated in fig. 71. AB is the object glass, of diameter d

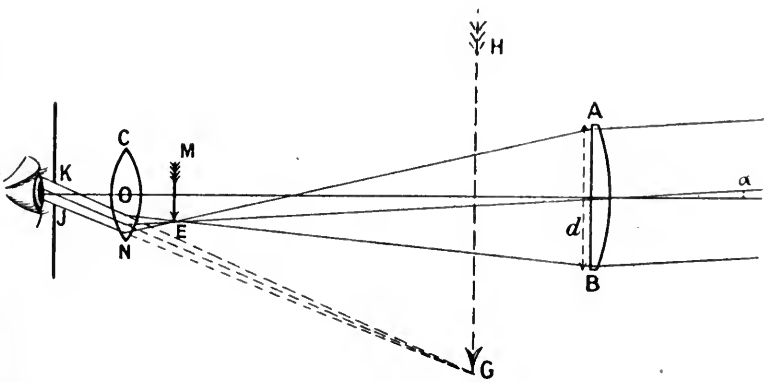


FIG. 71.

and focal length F, and CN is the eyepiece, which has a focal length f. AB forms a real and inverted image EM of the object in its focal plane; this image is just inside the focus of CN, which thus acts as a magnifying glass, producing a virtual magnified image of EM at GH. It is this virtual image at GH which is seen by the eye at KJ.

To find the magnification we can proceed as follows. Let 2α be the angle subtended by the distant object at the eye. Then, since the distance is great, 2α is also the angle subtended by the object at the object glass. Consequently the length of the image EM is $2\alpha F$. Since EM is just distant f from CN, the angle it subtends at O is $2\alpha F/f$. Since G is at a great distance OE is practically parallel to NG; this may be seen best from the graphical construction for determining G. Consequently $2\alpha F/f$ is also the angle subtended at the eye by the final image GH. Dividing the angle subtended by the final image by the angle subtended by the object we find therefore that the magnification is equal to F/f.

The same result may be obtained in a slightly different manner.

Let the object have length L and be situated at a distance u from the object glass. Then, since u is large, the angle subtended at the eye is L/u. Since image is to object in the ratio of their respective distances, $EM = LF/u$. Let D be the distance of GH from O. Then, applying the same rule again, $GH = (LFD)/(uf)$ and the angle subtended by GH at the eye is approximately $(LF)/(uf)$. Dividing this by the original angle subtended by the object we find again that the magnification is F/f.

Let the points J and K be respectively the images of the points A and B formed by the eyepiece. Let v be the distance of KJ from O. Then

$$\frac{1}{v} + \frac{1}{F + f} = \frac{1}{f},$$

if we use v, F, and f merely arithmetically,

$$v = \frac{f(F + f)}{F}$$

and

$$KJ = \frac{vd}{F + f} = \frac{fd}{F}.$$

The ray AE after refraction passes through J. All the rays which pass through both lenses pass through between K and J, and if a screen were placed there, it would show a bright circular disc. This disc, sometimes called the eye-ring, is according to Abbe's terminology the exit-pupil of the instrument. We see from the expression for KJ that the diameter of the object glass divided by the diameter of the eye-ring is equal to the magnification.

On leaving the instrument the rays come closest together at the eye-ring and that is the place for the eye. The instrument is usually designed with its eye-ring smaller than the eye-pupil, so that all the light from the object glass enters the eye. But to make fig. 71 clear it has been necessary to draw the eye-ring large.

Fig. 71 is intended merely to illustrate the theory of the astronomical telescope. In practice instead of a single object glass a combination corrected for chromatic aberration is used. In the case of small instruments, for example such as are used for spectroscopes, this combination very often takes the form described on p. 64, an equi-convex crown glass lens backed by a plano-concave flint glass lens. Also in practice instead of a single magnifying glass a Ramsden or Huygens ocular is employed.

The Ramsden eyepiece consists of two equal plano-convex lenses with their curved faces turned towards one another and the distance between them equal to two-thirds of the focal length of either. The Huygens eyepiece consists also of two convex lenses, but in it the lens farther from the eye has a greater focal length than the other, usually three times as great, and the distance between the lenses is twice the shorter focal length. In both eyepieces the lens next the eye is called the eye lens and the other is called the field lens. Denoting the focal length of the field lens by f and the focal length of the eye lens by f', the focal length of the combination is (cf. p. 46) given by

$$\frac{ff'}{f + f' + a} = \frac{3}{4}f$$

in the case of the Ramsden eyepiece, and by

$$\frac{3f'}{2}$$

in the case of the Huygens eyepiece.

Figs. 72 and 73 represent the positions of the lenses, principal

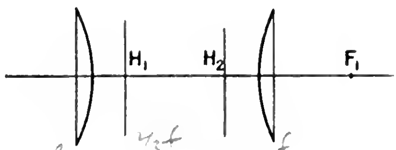


FIG. 72.

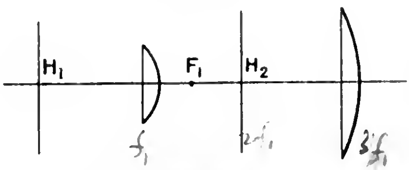


FIG. 73.

planes, and focal planes of the object space for these two eyepieces. The principal planes are crossed. It will be noticed that the focus of the Huygens eyepiece falls between the lenses; for this reason it is said to be negative, in contradistinction to the Ramsden eyepiece, which is said to be positive. Thus the Huygens eyepiece cannot be used to focus on cross-wires the way the Ramsden eyepiece can. Huygens' eyepiece was designed to diminish the effects of spherical aberration as much as possible.

The reason for using an eyepiece consisting of two lenses instead of a single magnifying glass is to diminish the chromatic aberration and other defects of the image as much as possible. They tend to be less owing to the refraction being distributed over four surfaces instead of two. The condition that the equivalent focal length should be the same for all colours (cf. p. 66) is fulfilled by the Huygens eyepiece and approximately fulfilled by the Ramsden eyepiece.

The performance of an eyepiece is made much better by the fact that it has to deal only with thin pencils of light. The pencil diverging from the point E of the image in fig. 71 is a thin one because it has had to come through the object glass of the telescope, and the ratio AB/EB is usually not more than 1/12. It is the same with rays diverging from the other points of EM.

An astronomical telescope can be made suitable for terrestrial objects by fitting it with an erecting eyepiece. Fig. 74 shows the

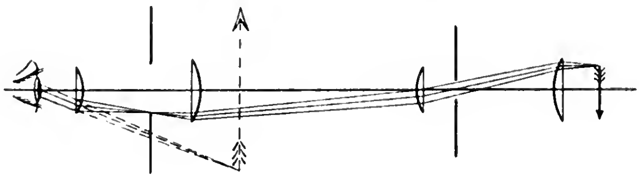


FIG. 74.

form of erecting eyepiece most used. It consists of four lenses and at the positions indicated diaphragms are fitted.

Magnification and Resolving Power of a Telescope. The magnification of a telescope is given by F/f. The diameter of the eye-ring $(fd)/F$ must not be greater than, at the outside, one-fifth of an inch, otherwise all the light does not enter the eye. If F, f, and d are subject only to this restriction, there is no limit to the magnification of a telescope. We find, however, that beyond a certain point nothing is gained by increasing the magnification. The image becomes larger but reveals no fresh detail. It is just like pulling out an elastic sheet on which a picture is painted.

If a telescope is focussed on a luminous point, a star for example, the image formed by the object glass is not a point, but a small disc surrounded by one or two faint concentric rings. The reason for this will be given afterwards in the chapter on diffraction; here it will only be stated as a fact. The radius of the dark ring immediately surrounding the disc is

$$1{\cdot}22\,\frac{F\lambda}{d},$$

where λ is the wave-length of the light used. If F is increased without a corresponding increase in d, the size of the disc increases. Every point of the object forms its own disc and thus the grain of the image increases.

The closest distance at which two stars can be recognised as separate is when the centre of the disc of the one falls on the innermost dark ring of the other, that is, when the distance between the images in the focal plane of the object glass is $1{\cdot}22\ F\lambda/d$ or when the angle between the stars is $1{\cdot}22\ \lambda/d$. This angle may be defined as the resolving power of the telescope. It is equal to

$$\frac{5{\cdot}0''}{d}$$

if the wave-length of light be taken as 2×10^{-5} inches and the diameter of the object glass be measured in inches. As a result of experience the astronomer Dawes gives $4{\cdot}5''/d$ instead of the above value.

The object glass of the Yerkes telescope, the most powerful one employed hitherto for astronomical observation, has a diameter of 40 inches and a focal length of 65 feet. Its resolving power is consequently about $\frac{1}{8}$ second. To grind and polish this object glass required great care and took a long time. It is also extremely difficult to obtain pieces of glass this size of the necessary homogeneity, and it is possible that if the object glass were larger it might strain under its own weight. So that the limit of resolving power in telescopes is probably nearly reached.

When d, F, and the maximum diameter of the eye-ring are given, the maximum value of f is given. Vision begins to be sensibly impaired when the diameter of the eye-ring is less than 1/30 inch, so that there is also a minimum value of f, about one-sixth of its maximum.

The resolving power of the eye according to Helmholtz lies between 1 and 2 minutes, so that in the case of the Yerkes object glass a magnification of about 720 would be required to bring out the full detail of the object. The maximum and minimum magnifications mentioned above are in this case 1200 and 200.

A telescope with a $1\frac{1}{4}$ inch object glass, such as is fitted to the larger spectrometers, has an object glass of focal length about 14 inches, and a Ramsden eyepiece, the focal length of each of the components of which is about $1\frac{1}{2}$ inches. Its magnifying power is therefore $14/\frac{9}{8} = 12\frac{4}{9}$ and its theoretical resolving power 4 seconds. The eyepiece is therefore scarcely powerful enough to utilise the full theoretical resolving power of the object glass, but is probably ample for its actual resolving power, as the object glass will be by no means perfect. As the apparent angular diameters of Venus, Jupiter, and Saturn vary respectively from 11 to 67 seconds, 32 to 50 seconds, and 14 to 20 seconds, such a telescope should show up their form.*

Apart from the actual measurement of the focal lengths of the lenses themselves, the magnification of a telescope can be determined very simply by two methods. The first method consists in illuminating the object glass by a lamp and receiving the image formed of it by the eyepiece, i.e. the eye-ring, on a ground glass screen. Then, if the diameter of this image is measured, the magnification is obtained by dividing it into the diameter of the object glass. Care should be taken, however, that it is really the image of the object glass rim that is obtained and not the image of some diaphragm inside the instrument. If there is any doubt in the matter a rectangular aperture should be used close up in front of the object glass. Then the image is rectangular and there is no possibility of mistake.

In the second method a white scale is fixed up at a great distance and a sliding mark placed on it. The telescope is then focussed on the scale and the latter is observed with one eye through the telescope and with the other eye direct. The two images superimpose. The

* One used by the Author shows four of Jupiter's moons and the phases of Venus, but makes nothing of Saturn's rings.

mark is then moved along the scale until the image of the part cut off by it, as seen through the telescope, is equal in length to the whole scale as seen direct. The magnification is then obtained by dividing the length of the whole scale by the length of the part cut off.

The simplest method of determining the resolving power of a telescope is to fix up a piece of wire gauze with some tissue paper behind it, the tissue paper being illuminated by a lamp, and then to find the greatest distance at which the wires of the gauze are separate when seen through the telescope. A simple calculation then gives the angle subtended by the distance between them.

Galileo's Telescope. The astronomical telescope was proposed by Kepler in 1611. A year or two previously Galileo had made and used a telescope of the type represented in the following figure.

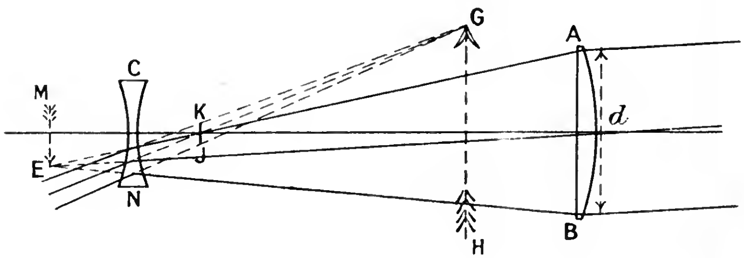


Fig. 75.

Atter passing through the object glass the rays from the distant object are converging towards a real inverted image, EM, when they fall upon a concave lens, CN, and form an erect and virtual image, GH, which is seen by the eye placed close up to CN. The image EM is just outside the focal length of CN.

If F and f denote respectively the focal lengths of the object glass and eyepiece, it may be shown in the same way as for the astronomical telescope that the magnification is F/f. Also the resolving power is given by the same formula as for the astronomical telescope. There are, however, some important differences.

First of all there is the great advantage for terrestrial use that the image is erect. Then the instrument is shorter than the astronomical telescope of the same magnifying power, its length being given by $F - f$ instead of $F + f$, where F and f stand for the numerical values of the quantities in question. Also it cannot be used with cross-wires. They would have to be placed at EM and would consequently get in the way of the eye. Again, the image of the object glass, which is a virtual one, is situated at KJ. Its size is fd/F and it is at KJ that the emergent rays come closest together. The section of the beam is much greater where it enters the eye, Consequently the field is much narrower than in the astronomical telescope with the same magnification, and its illumination falls off towards the edge.

Since the eyepiece is concave it may be made to correct part of the chromatic aberration of the object glass.

Ordinary opera-glasses consist of two Galilean telescopes mounted side by side and focussed by the same screw. The optical system consists usually of an object glass formed of a thin crown and flint lens and an eyepiece formed of a single concave lens, or the object glass and eyepiece may each consist of three lenses cemented together, each combination being achromatic by itself.

Prism Glasses. Two simple astronomical telescopes cannot be mounted together as an opera-glass on account of their length and the inversion of the image. In 1895 Messrs. Zeiss introduced their prism glasses, in which by the introduction of two right-angled prisms both

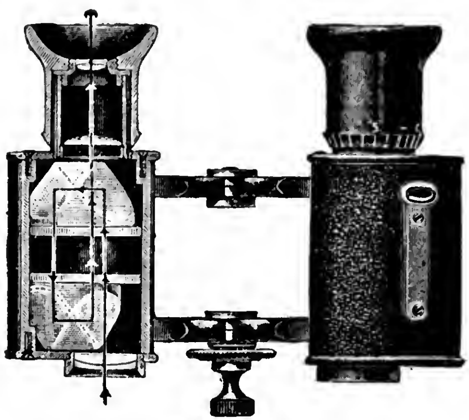


Fig. 76 (from R. and J. Beck's Catalogue).

these defects were remedied at once. The principle of the prism glasses had occurred to Porro in 1853, but they were not successful then, owing to want of uniformity in the glass available. Fig. 76 explains their construction. The rays after passing through the object glass traverse the body of the instrument once, are reflected by a right-angled prism which reverses the image right and left and then travel back to the second prism, which inverts the image and reflects them towards the eyepiece. The rays thus traverse the body of the instrument three times. Fig. 77 shows the image-reversing properties of a right-angled prism.

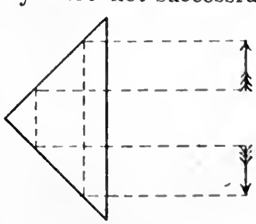


Fig. 77.

The advantage of prism glasses over the old form of opera-glasses lies in the wider field of view.

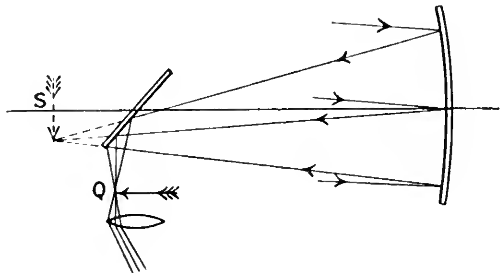
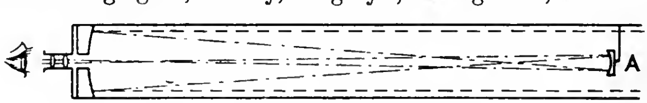


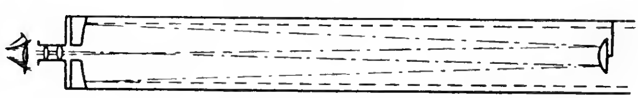
Fig. 78.

Mirror Telescopes. As a result of his discovery of the spectrum Newton learned the cause of chromatic aberration, and found that it was a much more serious error than spherical aberration in the telescopes of his day. As he believed it impossible to make an achromatic lens combination, he invented and constructed a telescope in which the object glass is replaced by a concave spherical mirror. Fig. 78 represents the principle of his instrument. The rays from the distant object are reflected by the mirror, and would form a real image at S were they not reflected by a right-angled prism or mirror so as to form the image at Q. This image is then examined by the eyepiece.

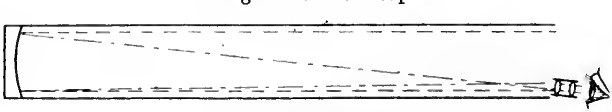
The principles of three other types of reflecting telescope are shown in the following figure, namely, Gregory's, Cassegrain's, and Herschel's.



Gregorian Telescope.



Cassegrainian Telescope.



Herschelian Telescope.

Fig. 79.

In Gregory's telescope the rays after reflection from the large reflector fall on the concave mirror and form a real image which is examined by the eyepiece through a hole in the large mirror. Cassegrain's telescope differs from Gregory's only in having a convex mirror instead of a concave mirror at A. In Herschel's telescope the large mirror is slightly inclined, the secondary mirror is dispensed with, and the observer stands with his back to the object. His head, of course, partially obstructs the light, so that the arrangement is only practical with very large instruments.

Mirrors for reflecting telescopes were originally made of speculum metal, a somewhat brittle alloy of copper and tin, and when the surface tarnished, it had to be repolished, and thus the most difficult and critical part of its construction repeated. They are now made of glass and silvered, and an old silver film can quite easily be replaced by a new one.

Large reflecting telescopes are, of course, considerably less expensive than refracting instruments of the same power, but they are inconstant and require careful attention, and so to-day the refractor is regarded as superior, although for the century and a half after Newton the superiority of the reflector was unquestioned.

The Microscope. If the focal length of a simple magnifying glass is diminished in order to obtain increased magnification, the eye has to be placed inconveniently close to the object examined ; also lenses of such small focus are difficult to grind accurately. The requirements as to freedom from aberration are also too great for a single lens to satisfy. For large magnifications it is therefore necessary to use a compound microscope. The principle of the latter instrument is represented in fig. 80. In the compound microscope the work of forming the image is distributed over two lenses or lens systems, the object glass and the eyepiece. It is the function of the object glass to form an image of the object by as wide angle a pencil as possible, and it is the function of the eyepiece to form a large image of this image by means of thin pencils.

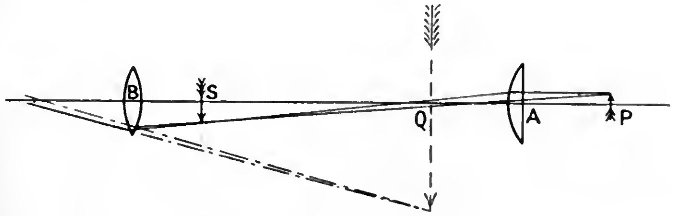


FIG. 80.

In fig. 80, which is purely illustrative, P is the object, A the object glass, S the image formed by the object glass, B the eyepiece, and Q the enlarged virtual image of S formed by B. Let F and f be the numerical values of the focal lengths of the object glass and eyepiece. Let the image Q be situated at the distance of distinct vision, D, from the eye, which is the same thing approximately as the distance from the lens B. Let v be the distance of S from the lens A and let L be the length of the object.

The object is situated just beyond the focus of A so the length of S is approximately $(v\text{L})/\text{F}$. S is just inside the focus of B so that the length of Q is approximately $(v\text{LD})/(\text{F}f)$. Hence the magnification is

$$\frac{v\text{D}}{\text{F}f}.$$

Microscopy is not a part of physics but a science in itself. Microscopes are used extensively in medicine and the biological sciences, also in petrology for examining sections of rocks. The object examined is placed on a glass slide and illuminated by transmitted light. The sections of opaque objects taken are so thin as to be translucent. Below the stage of the instrument is usually placed an arrangement of lenses called a condenser for the purpose of illuminating the object.

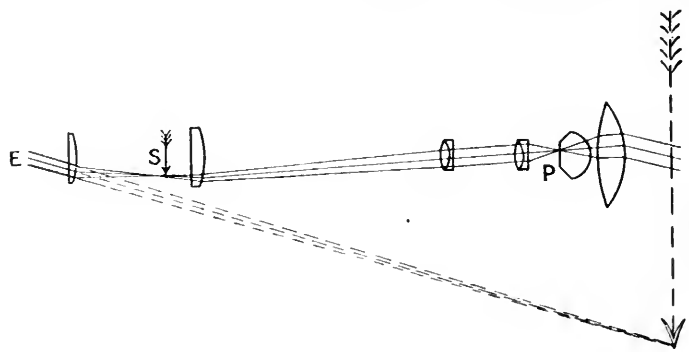


FIG. 81.

Fig. 81 is a sketch of the optical system of a typical microscope with a low-power object glass in. The light comes from the right from the source, a flame for example, and passing through the two lenses of the condenser is made to converge on P, the object to be examined. It diverges from P, passes through the two achromatic lenses of the object glass and the field lens of the eyepiece, and forms a real image at S. It then passes through the eye-lens and enters the eye at E appearing to come from the large virtual image. For clearness only the rays diverging from one point of the object are shown.

The eyepiece is usually of the Huygenian type, and at S is placed a diaphragm limiting the field of view. The eye is placed at the eye-ring where the rays come closest together. The object glass must be corrected for chromatic aberration and spherical aberration and must also obey the sine condition. The importance of this last condition was first realised theoretically by Abbe, but the good objectives that had been made previously obeyed it.

§ To find the resolving power of the object glass let us go back to our former result for the telescope object glass. The latter received parallel rays and formed an image in its focal plane. In the case of the microscope objective the conditions are reversed; the object is near the focal plane and the rays after passing through the objective are nearly parallel.

Two points in the image formed by a telescope objective can just be separated when the distance between them is $1{\cdot}22\ \mathrm{F}\lambda/d$. If $2a$ is the angle subtended at the image by a diameter of the objective, $\tan a = d/2\mathrm{F}$. The distance between the two points in the image can thus be written

$$\frac{\cdot 61\lambda}{\tan a},$$

and if they were object instead of image we should expect them to be resolved by the objective. The expression above might thus give approximately the least distance between two lines that can just be separated by a microscope objective.

The angle a is so much greater in the case of the microscope, though, that the one case cannot be derived from the other, and the matter is also complicated by the fact that the points resolved by a microscope are not independent self-luminous objects. They are both illuminated by the condenser from the same part of a flame, and hence the light from them is in a condition to interfere. Even at critical illumination, i.e. when an image of the flame is formed on the object by the condenser, this image is never sharp enough to make the light transmitted by two points so close together quite independent.

The fact that the points to be resolved are not self-luminous was first taken into consideration by Abbe and illustrated by some striking experiments. He was led to the result that the closest distance that could be separated was

$$\frac{\lambda}{2\mu \sin a},$$

where μ was the index of refraction of the medium between object and objective. The light from the object must of course fill the whole aperture of the objective. The numerical value of the above result is not greatly different from the expression higher up on the page. To the product $\mu \sin a$ Abbe gave the name numerical aperture (N.A.).

If P represents the object and APB the cone of light falling on it from the condenser, P causes diffraction figures in the diverging cone, and the detail shown depends on the number of these figures that enters the object glass L, i.e. it depends on the aperture of the wave-front used. If the space between P and the object glass is filled with oil of index of refraction μ, owing to the refraction of the rays, the cone CPD is filled with light that originally filled the cone A′PB′, an incident wave-front of greater aperture is used, and hence the greater resolving power. With such an immersion system there is also a

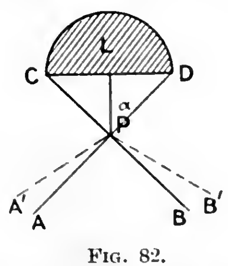


FIG. 82.

gain of light, both owing to the wider cone and to the elimination of reflection losses. The greatest value of the N.A. obtainable is about 1·6, so if λ is taken as $5{\cdot}3\ 10^{-5}$ cms., the limit of microscopic resolution is

$$\frac{\lambda}{2\mu \sin \alpha} = \frac{5{\cdot}3\ 10^{-5}}{2 \times 1{\cdot}6} = 1{\cdot}7\ 10^{-5} \text{ cms.}$$

This of course presupposes a sufficiently perfect optical system, i.e. the detail gained by the high aperture must not be lost through chromatic or spherical aberration, etc.

The least angle that can be separated by the normal eye is about 1 or 2 minutes, say 1·5 minute. At the distance of distinct vision, 25 cms., this angle is subtended by $1{\cdot}1\ 10^{-2}$ cms. Hence the magnification necessary for the above resolving power is

$$\frac{1{\cdot}1\ 10^{-2}}{1{\cdot}7\ 10^{-5}} = 650 \text{ approximately.}$$

Increasing the magnification beyond this will bring out no more detail. Sets of gratings * with the distance between the rulings increasing in graded steps are used for testing the resolving power of microscopes.

To measure the magnifying power of a microscope a fine scale with divisions of known length is placed upon the stage, and the eye focusses on it and at the same time looks at another scale outside the microscope and placed at the distance of distinct vision. The simplest way of doing this is by means of a little piece of clear glass fixed at 45° immediately above the eyepiece, so that the rays from the eyepiece pass through it directly and the rays from the other scale are reflected from the side. The two scales are seen superimposed, and it is then easy to say how much the divisions on the one have been magnified.

§ There is much accessory apparatus for use with the microscope. For example, the camera lucida is an arrangement in principle similar to that described above for obtaining the magnification, by which the magnified image is seen superimposed on a drawing-board and can be sketched. The ordinary eyepiece can be replaced by a spectroscopic one in which there is a slit in the plane of the image S (fig. 81) and a prism between the eye and the eye-lens. For crystallographic work a polarising prism is fitted below the stage, and an analysing prism either immediately above the objective or above the eyepiece.

In photomicrography, i.e. photographing the magnified image produced by a microscope, the tube of the microscope is placed horizontal as the camera requires a long extension. The tube of the microscope is also kept horizontal in projection work.

When a very thin intense beam of parallel light is passed horizont-

* Sold under the name of Grayson's rulings. The one at 5s. giving 1000 lines to the cm. is a very convenient test object for students to use with a vernier microscope. The aperture can be stopped down until the lines are not separated.

ally into certain colloidal solutions, the particles in the solution in the path of the beam scatter the light, and, if they be focussed on from above by a microscope the tube of which is vertical, they can be seen by the scattered light. This arrangement is known as the ultra-microscope. By it particles can be seen which are far too small to be made visible by the ordinary means of illumination: indeed we can see down to those which have diameters of $10\mu\mu$ or $\frac{1}{60}$th of the wave-length of Na light. Like the stars their angular dimensions are too small to give them any magnitude: they are simply point sources of light.

Photographic Camera. A photographic camera consists essentially of a rectangular box at one end of which is placed the sensitive plate or film and at the middle of the opposite end of which is placed the lens. The action of the photographic plate will be explained later. The distance of the lens from the plate can be altered so as to focus the picture sharply, and no light must enter the camera except through the lens. By means of a series of stops placed at the lens the aperture of the pencils of light forming the image can be regulated. The diameter of the aperture is always expressed as a fraction of the focal length of the lens; thus, if a lens is working at *f*/16, we mean that the diameter of the stop is $\frac{1}{16}$th of its focal length. A lens working at *f*/8 thus admits about 8 times as much light as one working at *f*/22.

The ordinary dry plate is sensitive between $\lambda = 2{\cdot}2$ and $\lambda = 5{\cdot}0\ 10^{-5}$ cms., but the glass of the lens does not permit the light between $\lambda = 2{\cdot}2$ and $\lambda = 3{\cdot}3\ 10^{-5}$ cms. to pass. The plate is most sensitive to the violet. The lenses must be achromatised for this range of wave-lengths and not for visual use.

The photographic lens has as a rule to include a wide angle of view in a picture on a flat surface, and consequently astigmatism, curvature, and distortion are much more serious defects than in the telescope and microscope. All defects except distortion can be diminished at the expense of the brightness of the image by stopping down the lens.

Fig. 83 represents two well-known types of lens, the Petzval portrait lens and the rapid rectilinear lens. The Petzval portrait lens

FIG. 83.

is extremely rapid, working at *f*/4, and besides portrait work is used as projecting lens with the optical lantern. It has the stop in the middle and consists of a cemented achromatic pair as the first element with a back pair separated by a small interval. It possesses distortion, however, and the illumination falls away towards the edge of the plate.

The rapid rectilinear lens is symmetrical, consisting of two achromatic components made of a convex crown meniscus and concave flint meniscus. The stop is placed midway between the two components. The lens is orthoscopic or free from distortion, the one component producing barrel-shaped distortion, i.e. making the sides of the image of a square convex outwards, and the other component producing cushion-shaped distortion, i.e. making the sides of the image of a square concave outwards. So together they neutralise one another.

For a long time rapid rectilinear lenses were extremely popular but they are now outclassed by the anastigmat.

Telephotography. The size of the picture of a distant object varies as the focal length of the lens. But if a long focus convex lens is used it requires to be placed at a great distance from the plate and thus the camera is inconveniently long. A telephoto lens combination obviates this difficulty and enables us to take highly magnified pictures with an ordinary camera. It consists of a convex and a concave lens, both achromatic, with the concave lens near the position where the ordinary lens of the camera is usually placed, and the convex lens

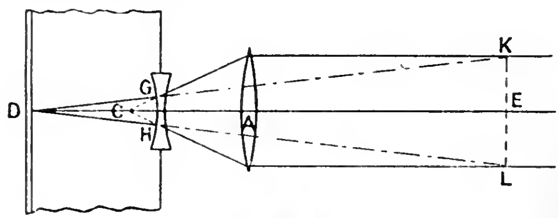


Fig. 84.

outside the camera in front of the concave lens. Fig. 84 illustrates the arrangement. The camera lens is removed; GH is the concave lens and A is the convex lens. A beam of parallel light is shown incident on A. It would come to a focus at C were it not for the concave lens which makes it less convergent and brings it to a focus at D. By producing the rays DG and DH backwards to cut their original directions at K and L we find that KL is the principal plane of the image space and DE is the equivalent focal length of the instrument. By increasing the distance between the lenses the principal plane is moved further out and consequently the magnification increased.

Optical Lantern. Fig. 85 depicts the optical system of a lantern used for projecting slides. AB represents the slide; it is placed upside down. C is the objective or projecting lens. It forms an erect image of the slide on the screen, and by focussing it the image can be made sharp for different positions of the screen. The source of light, E, is an electric arc, the positive carbon of which is horizontal and the negative vertical. By this arrangement the full light from the crater

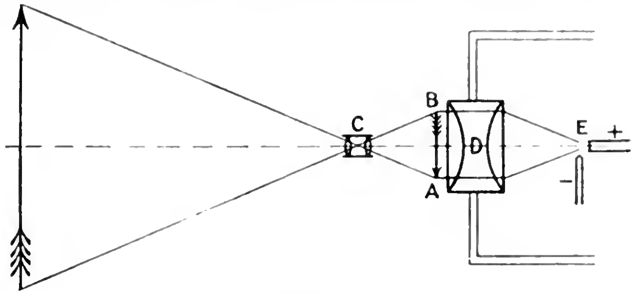


FIG. 85.

falls upon the condenser, D. The latter consists usually of two plano-convex lenses with their plane faces outwards, and its function is to make as much light as possible pass through the slide in such a direction that it also passes through the objective. The electric arc is enclosed in a box so as to keep in all light except that which forms the image on the screen. Lime-light may be used instead of the arc, but the arc is brighter and more convenient and is always to be preferred if available.

Under the name of epidiascopes there are different arrangements now sold for projecting opaque objects. They all require very powerful light sources. Fig. 86 shows how an ordinary lantern may be adapted for the purpose. M and N are two mirrors. M throws the light from the condenser on to the opaque object, AB, which then acts as a new source and sends out rays, which, after reflection by the mirror N, are focussed by the objective on the screen. In the epidiascope the object and mirrors must be very carefully covered in, because owing to the fainter image stray light is more important than in the ordinary use of the lantern for projecting slides.

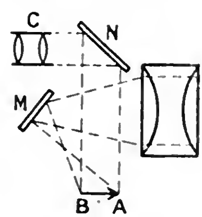


FIG. 86.

Visual impressions on the retina persist after the removal of the stimulus for about $\frac{1}{16}$th of a second. Thus, if photographs are taken of a moving object at a rate of not less than 16 per second, and if these photographs are projected on a screen at the same rate, the discontinuous pictures fuse together and produce an illusion of continuous motion. This is the principle of the cinematograph. The pictures are smaller than lantern slides, the regulation size being about 1 inch by ¾ inch, and they are projected by a lens of about 2¾ inches equivalent focal length. They are printed on a film of celluloid or a similar preparation, and the film is uncoiled from one spool, is jerked by a mechanism through the "gate" in the focal plane of the projecting lens, and coiled up on another spool. Each jerk moves the film a distance equal to the height of a picture. Each picture rests for an instant while in the gate and is then projected, and, while it is being jerked out and the next one jerked in, a sector passes up in front of the projecting lens and cuts off the light. The films are tough enough to stand the strain of being pulled through but the celluloid ones are inflammable, and the image of the crater of the arc on one for a short time is sufficient to set it on fire.

The Sextant. It has been shown that when a mirror is rotated the angle turned through by the reflected ray is twice the angle turned through by the mirror. The sextant, which is founded on this principle, is an instrument used for measuring the angle subtended at the

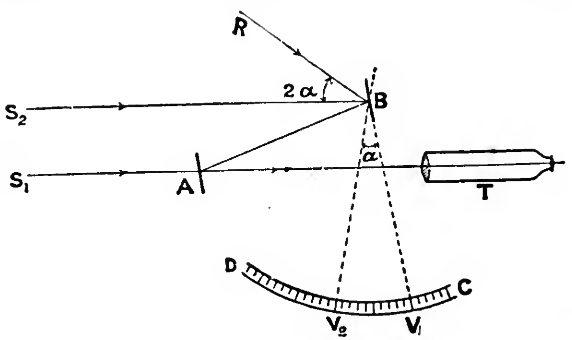


FIG. 87.

observer by two distant objects. It consists of a telescope, T, directed towards a mirror, A, only half of which is silvered and the rest transparent, a mirror, B, which can be rotated, and a pointer attached to B, which reads its position on a graduated circle CD.

Consider an object at such a distance that S_1 and S_2, the rays from it to the mirrors A and B, are parallel. Let S_1 be the ray from it to the mirror A, and let the mirror B be in such a position that the ray S_2 after reflection at B and A enters the telescope in the direction S_1. Then the two images of the object will be seen superimposed. The reading of the pointer for this position of B is taken as zero. Suppose that, starting from this zero and keeping the telescope directed on one of the objects, it is necessary to rotate B through an angle a to bring the images of two different objects into coincidence. Then the angle they subtend at the eye is $2a$. The angles on the scale CD are numbered at twice their proper value. Hence the reading V_1V_2 of the pointer gives the angle subtended directly.

The sextant is used principally for taking the angle of elevation of the sun. The arrangement usually adopted on land is represented in the diagram. The observer sits in front of a mercury trough, resting the elbow of the arm that holds the instrument on the knee. The telescope is directed towards the image of the sun in the mercury, and the image of the sun formed by reflection on the mirror is brought to coincide with it by moving the pointer along the graduated circle. The scale reading then gives $2a$ (cf. fig. 88), and since AB and CD are parallel, a is the altitude of the sun. A mercury surface is used instead of a mirror because it sets itself horizontally under the action of gravity.

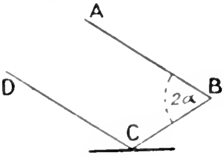


FIG. 88.

At sea the angle between the sun and the horizon is taken.

It is easy to see from a figure, since the position of the sun in the heavens is known, that the latitude can be determined from its altitude at noon. If the time of its highest altitude is noted on the chronometer, which keeps Greenwich time, the time the earth has been turning since the sun was above the meridian at Greenwich is known, and hence the longitude can be determined.

EXAMPLES.

(1) If half the object glass of a telescope which is pointed at the moon is covered, how will the appearance of the moon as seen through the telescope be affected?

(2) What is theoretically the angular distance between the centres of two stars which are just separated by a telescope with a 9-inch object glass?

(3) Rule two lines close together on a card and look at them through a magnifying glass held close up to one eye. At the same time look with the other eye at a scale placed at the distance of distinct vision. By getting the image of the lines to superimpose on the scale and measuring the distance between the lines the magnification can be determined. Compare it with the theoretical value.

(4) Determine the magnification of a spectroscope telescope (a) by taking it to pieces and measuring the focal lengths of the different lenses separately, then using the formula F/f; (b) by the method of the illuminated scale with the sliding mark described on p. 85; and (c) by measuring the diameters of the object glass and eye-ring as described on the same page.

(5) Determine the resolving power of the same telescope by fixing up a piece of wire gauze with illuminated tissue paper behind it and finding the greatest distance at which the wires are seen separate. Compare the result with the theoretical value. Stop down the object glass of the telescope and note the change in the experimental value.

(6) Look at Venus, Jupiter, and the moon with the same telescope.

(7) Measure the magnification of a microscope by the method described on p. 92, then compare the result with the value given by the theoretical formula.

CHAPTER VII.

THE SPECTROMETER AND THE DETERMINATION OF INDICES OF REFRACTION.

THE most straightforward way of determining the index of refraction of glass is by using it in the form of a prism with the spectrometer. A prism in geometry is a polyhedron which has two of its faces parallel, equal and similar polygons, and the other faces of which are parallelograms, but in optics it always means a right prism on a triangular base. Any plane perpendicular to the sides of the prism is called a principal plane. Obviously, if a ray of light is incident on one of the sides in a principal plane, it remains in that plane during its passage through the prism and after leaving it.

Let fig. 89 represent the section of a glass prism by a principal plane, and let PQ be a ray of light incident on the face AB at Q. After refraction at Q it travels in the direction QR, and after refraction at the

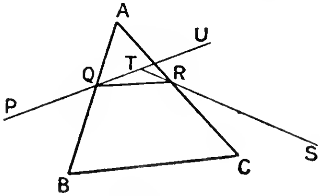


FIG. 89.

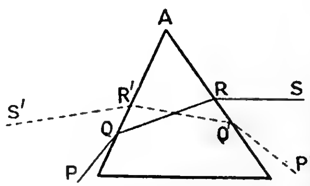


FIG. 90.

second face at R it emerges in the direction RS. The edge A, i.e. the edge in which the faces that the light passes through meet, is termed the refracting edge of the prism. Produce PQ to U and produce SR to meet it in T; then the angle UTS between the directions of the incident and emergent rays is the deviation produced by the prism.

If the angle of incidence of the ray PQ on the face AB alters, the deviation alters. If the incident and emergent rays are equally inclined to the prism, i.e. if AQ = AR, the deviation is a minimum, and the prism is said to be in the position of minimum deviation. This is easily seen. For, let the deviation of any other ray PQRS (fig. 90) be a minimum; then the deviation of the ray P'Q'R'S' which passes through the prism in the symmetrically opposite direction, i.e. so that AQ = AQ' and AR = AR', is the same as the deviation of PQRS. Consequently the latter cannot be a minimum.

Suppose now that fig. 91 represents the position of minimum

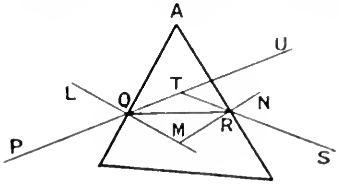


FIG. 91.

deviation, that LM and MN are the normals at the points of incidence and emergence of a ray PQRS. Write

$$\angle \mathsf{PQL} = \angle \mathsf{SRN} = i \quad \text{and } \angle \mathsf{RQM} = \angle \mathsf{QRM} = r$$

and let the angle of minimum deviation be δ. Then

$$\begin{aligned}\delta &= \angle \mathsf{UTS} = \angle \mathsf{TQR} + \angle \mathsf{TRQ}\\ &= \angle \mathsf{TQM} - \angle \mathsf{RQM} + \angle \mathsf{TRM} - \angle \mathsf{QRM}\\ &= 2(i - r).\end{aligned}$$

Also $\angle \mathsf{AQR} + \angle \mathsf{ARQ} = \pi - \mathsf{A}$

and $\angle \mathsf{AQR} + \angle \mathsf{ARQ} = \angle \mathsf{AQM} - \angle \mathsf{RQM} + \angle \mathsf{ARM} - \angle \mathsf{QRM}$

$$= \pi - 2r.$$

Hence $\mathsf{A} = 2r$, i.e. $r = \mathsf{A}/2$, and on substitution in the expression above for δ

$$\delta = 2\left(i - \frac{\mathsf{A}}{2}\right) \text{ or } i = \frac{\mathsf{A} + \delta}{2}.$$

But the index of refraction, μ, is defined by

$$\mu = \frac{\sin i}{\sin r}.$$

On substitution for i and r this gives

$$\mu = \frac{\sin \frac{\mathsf{A} + \delta}{2}}{\sin \frac{\mathsf{A}}{2}}.$$

Hence, if A and δ are known, μ can be calculated. A useful way of remembering this formula is by considering what happens if the prism is made of air instead of glass—δ becomes 0 and μ is equal to 1.

The spectrometer or goniometer is an instrument for determining A and δ. Fig. 92 shows a plan of one, arranged for determining δ. It consists essentially of a divided circle ABC about the axis of which a collimator, DE, and a telescope, FG, can rotate. The collimator is a tube at one end of which, E, there is an achromatic convex lens, and at the other of which, D, there is a slit exactly at the focus of this convex lens. The rays of light which enter the slit thus form a parallel pencil after passing through the lens. After deviation by the prism they fall on the object glass of the telescope, which is an astronomical one, and converge to form a real image, which is observed through the Ramsden eyepiece. The prism rests on a table which can be rotated about the axis of the divided circle and the rota-

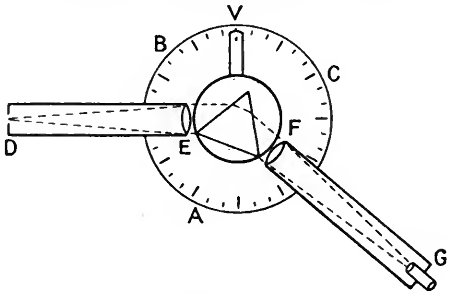


FIG. 92.

tion of which can be read by a vernier at V. The position of the telescope on the divided circle can also be read by a vernier. As δ varies with the colour of the light, it is usual to employ as light source the bright yellow monochromatic light produced by heating a sodium salt in a Bunsen burner. The best salt to use is the bicarbonate: a bead of this salt formed in a small loop at the end of a thin platinum wire and held in the edge of the flame gives an intense yellow colour for a long time.

To measure δ the prism is first removed and the telescope directed to look into the collimator. An image of the slit is then seen in the field and the telescope is turned until this image coincides with the cross-wires. The reading is then taken. The prism is next put into position on the prism table and the telescope turned so as to see the image formed by the light which passes through the prism. The prism is then rotated so as to bring it into the position of minimum deviation. The latter is easily recognised, because at it the image turns in the field and begins to move in the other direction. When it is found the reading on the circle is again taken. The difference of the two readings gives δ.

To measure A, the angle of the prism, we may proceed in either of two ways. In the first of these (fig. 93) the prism is placed with its edge A near the centre of the prism table so that the rays from the collimator are incident on it in the direction PA, some on the face AB, and some on the face AC. The rays are reflected in the directions AQ and AR. If the prism is kept fixed and the telescope pointed in the directions QA and RA, images of the slit will be seen. Hence $\angle QAR$ can be measured. The angle of the prism, A, is half $\angle QAR$. For, draw AM and AN the normals to the two faces. Then

$$\angle QAR = 2\pi - \angle QAP - \angle RAP = 2\pi - 2\angle PAM - 2\angle PAN$$
$$= 2(\pi - \angle MAN) = 2A.$$

In the second method the prism is adjusted so that the rays reflected from one of the faces produce an image of the slit coinciding with the cross-wires in the field of view. The telescope is then kept fixed and the prism table rotated until an image of the slit formed by

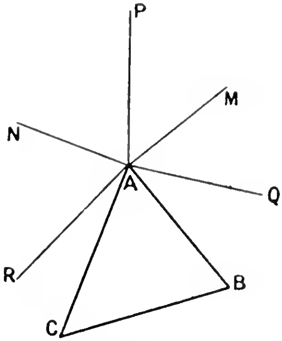


FIG. 93.

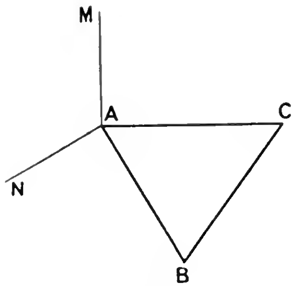


FIG. 94.

reflection from one of the other faces is coincident with the cross-wires. The difference of the angle of rotation and two right angles is equal to the angle between the faces from which the light was reflected.

For it is clear (fig. 94) that in the second position the normal to the one face has the same direction as the normal to the other in the first position. The angle turned through by the prism is hence equal to $\angle MAN = \pi - A$.

Figs. 95 and 96 represent two types of spectrometer which are in

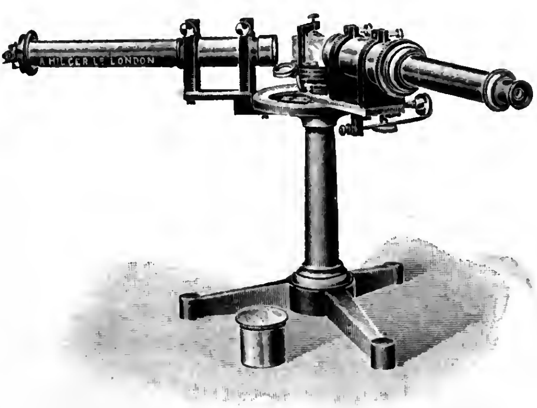


FIG. 95.

much use. In the second, which is the more elaborate of the two, the telescope is counterpoised so as to avoid strain during its rotation, there is a clamp and slow motion for both the prism table and the

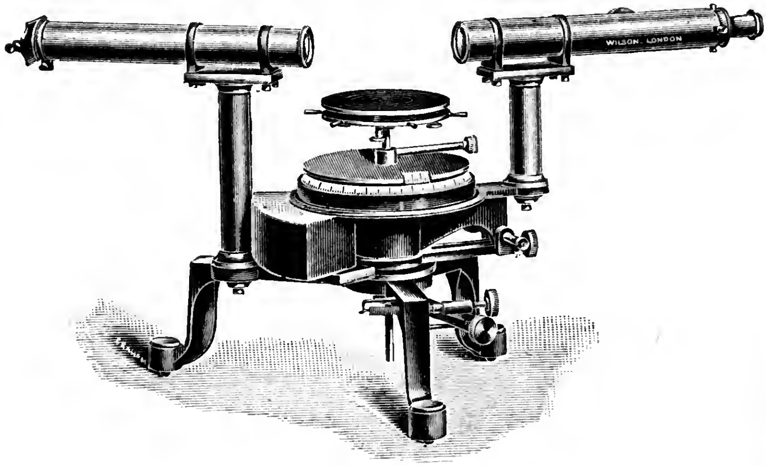


Fig. 96.

telescope, and the prism table can be raised and lowered and is provided with levelling screws.

Adjustments of the Spectrometer. It is necessary in working with the spectrometer that

(*a*) The telescope should be focussed for parallel light.

(*b*) The collimator should be focussed for parallel light.

(*c*) The optical axes of the telescope and collimator should be perpendicular to the axis of rotation of the instrument.

(*d*) The refracting edge of the prism should be parallel to the axis of rotation of the instrument.

The simplest way to make the adjustment (*a*) is to remove the telescope from the instrument, and pointing it at a white surface to adjust the eyepiece, until the cross-wires are as sharp as possible when seen through it. This fixes the relative distance of the eyepiece and cross-wires. Then the telescope is taken to an open window, directed towards some distant object, such as a church spire or chimney stalk, and focussed by means of the rack and pinion until the image of this object is as sharp as possible. This fixes the relative distance of the object glass and cross-wires and completes the adjustment.

The telescope is next replaced in the spectrometer and turned to look into the collimator. The slit of the latter is then illuminated with sodium light and its distance from the object glass adjusted, until its image seen through the telescope is as sharp as possible. This focusses the collimator for parallel light.

If the telescope cannot be removed from the instrument as in the case of the spectrometer represented in fig. 96, after focussing the eyepiece on the cross-wires the best way of proceeding is by the following method due to Schuster. The slit is illuminated with sodium light and the prism placed on the prism table so that the deviation of the refracted image is greater than minimum deviation. Then, for a given deviation there are two possible positions of the prism, one represented (fig. 97) by the full lines and giving a broad image of the slit and the other represented by the dotted lines and giving a narrow image of the slit. Let us suppose that the prism is in the position represented by the dotted lines, i.e. the rays from the collimator fall on it more obliquely than for minimum deviation. Also let the telescope and collimator be out of adjustment and the image be blurred. Focus the telescope till it appears sharp. Then rotate the prism into the other position. The image becomes blurred again. This time focus the collimator until it is sharp. Next rotate the prism into its first position, and if the image is not quite sharp, make it so by focussing the telescope again. And so on. When the image is sharp in both positions, the telescope and collimator are focussed for parallel light. In practice it is usually not necessary to focus more than three times; also if a mistake is made and the telescope and collimator focussed in the wrong order, this is at once indicated by the adjustment becoming rapidly worse.

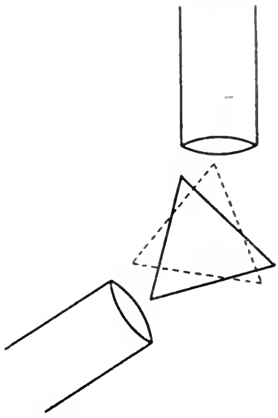


FIG. 97.

The principle of this method will be grasped readily from fig. 98.

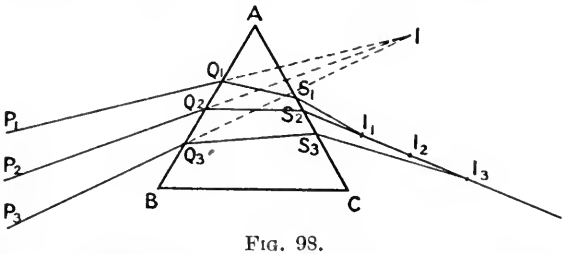


FIG. 98.

P_1IP_3 is a pencil of light which would converge to a point I if it were not for the prism ABC. The prism is placed so that the ray P_2Q_2 suffers minimum deviation. Let us suppose that a very thin pencil, of which P_2Q_2 is the principal ray, converges to I_2. Then the ray P_1Q_1 after refraction must cut S_2I_2 at a point I_1 between S_2 and I_2, and the ray P_3Q_3 after refraction must cut S_2I_2 at a point I_3 beyond I_2. If we regard P_1IP_2 as a separate pencil therefore, it converges more after refraction, and if we regard P_2IP_3 as a separate pencil, it converges less after refraction. That is, if the pencil is incident less obliquely than for minimum deviation it becomes less parallel, and if it is incident more obliquely than for minimum deviation it becomes more parallel after refraction by the prism. Similarly it may be shown by drawing another figure that the same holds true in the case of a pencil diverging from a point.

Now consider the practical case. Suppose that the telescope was in focus when the incidence of the light was less oblique than for minimum deviation, and that the prism is turned into the other position. The rays entering the telescope become more parallel, and refocussing the telescope obviously improves it for parallel light. Now turn the prism back into the first position. The pencil entering the telescope becomes too convergent or divergent as the case may be, and if we correct this with the collimator, we obviously bring the latter nearer its adjustment for parallel light.

The adjustment (*c*) requires to be made very seldom. In the simpler instruments it is done roughly once for all by the makers. In the case of the more accurate instruments the universal way of doing it is by means of a Gauss eyepiece.

The Gauss eyepiece is represented in fig. 99. It is merely a Ramsden eyepiece with an opening in the side of the tube and a plate of clear glass, B, between the lenses inclined at 45° to the axis of the tube.

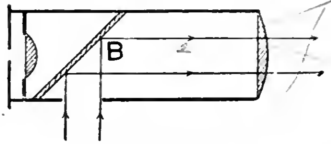


FIG. 99 (from Watson's "Practical Physics").

To set the optical axis of the telescope at right angles to the axis of rotation of the instrument, fix up a plane parallel plate of clear glass on the prism table as parallel to the axis of rotation of the instrument as can be done by the eye. Direct a beam of light through the opening, as shown in fig. 99, so as to be reflected by the glass and pass down the telescope and through the object glass. Turn the prism table so that the plate of glass mounted on it reflects this beam back into the telescope, through the eyepiece, and into the eye of the observer. The latter should then see an image of the cross-wires alongside the cross-wires themselves. If he does not see this at first, he will probably see an image of part of the circular edge of the aperture on which the cross-wires are mounted. This image can be made sharp with the rack and pinion—this is equivalent to focussing the telescope for parallel light—and then the image of the cross-wires will become visible. Next rotate the prism table through 180° without disturbing the glass plate on it. An image of the cross-wires will again appear but in a different part of the field. This is owing to the glass plate not being exactly parallel to the axis of rotation and should be remedied by the levelling screws of the prism table. Then, when the image appears at the same place in the field both times, the inclination of the axis of the telescope should be altered by whatever arrangement is provided for the purpose until the image coincides with the cross-wires themselves. The axis of the telescope is then at right angles to the axis of rotation of the instrument. Of course, if the faces of the plate of glass employed are not exactly parallel, two images of the cross-wires are formed, one by each face, but these images are not far enough apart to cause trouble.

When the axis of the telescope has been adjusted, the inclination of the collimator should be altered until the cross-wires appear half-way up the image of the slit. The axis of the collimator is then at right angles to the axis of rotation of the instrument.

There are two methods of making the adjustment (*d*). Let us suppose that the edge A is to be made parallel to the axis of rotation. The first thing to do is to set the prism so that one of the faces meeting in A, AC, say, is at right angles to the line joining two of the levelling screws, R and Q. Then there are two ways of proceeding. We may either use the Gauss eyepiece, first of all setting the telescope perpendicular to the face AC and using all three levelling screws to get coincidence of the cross-wires and their image, then setting the telescope perpendicular to the face AB and obtaining coincidence of the cross-wires and their image by the use of the screw P alone. It does not disturb the adjustment of the other face. Or we may direct the collimator towards the edge A and illuminate the slit. Then images of the slit are formed by reflection from both the faces AC and AB. These are examined with the telescope and the levelling screws of the prism table adjusted so that both these images appear at the proper height in the field, i.e. are bisected by the cross-wires. The image formed by the face AC is adjusted first by means of all three screws and then the image formed by AB by means of P alone.

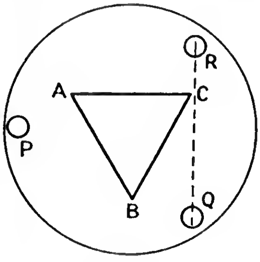


FIG. 100.

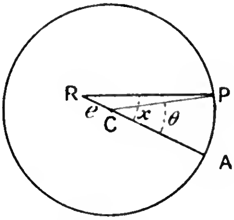


FIG. 101.

Both these methods assume that the telescope is at right angles to the axis of rotation of the instrument, and the second assumes that the collimator is also at right angles to the axis of rotation. The second is the easier of the two.

In accurate spectrometers there are two verniers for reading the rotation of the telescope, one exactly 180° round from the other, and similarly there are two verniers for reading the rotation of the prism table. It is usual to take the degrees from the one vernier and to take the mean of the minutes and seconds from both. This eliminates error due to eccentricity of the divided circle, that is, owing to the axis of rotation not passing through the geometrical centre of the divided circle. For, let C be the geometrical centre, R the centre of rotation, and P the position of, for example, the telescope arm. Let RC = e, produce RC to meet the circumference in A and let CP = r. Then $\theta = \angle \text{PCA}$ gives the apparent position of the telescope and $x = \angle \text{PRC}$ its true position. Then in $\triangle$RCP

$$\frac{e}{r} = \frac{\text{RC}}{\text{CP}} = \frac{\sin \text{RPC}}{\sin \text{CRP}} = \frac{\sin (\theta - x)}{\sin x}$$

or, since $\theta - x$ is small,

$$\frac{e}{r} \sin x = \theta - x.$$

Hence

$$x = \theta - \frac{e}{r} \sin x.$$

If RP rotates through 180° and x is increased by 180°, $\sin x$ has the same numerical value but changes sign; hence the difference between the apparent and true value of the angle changes sign and cuts out if the mean of the two readings is taken.

The Abbe or Auto-collimating Spectrometer. The essential feature of this instrument is, that the telescope performs the functions of both telescope and collimator. There is a slit in the plane on which the eyepiece focusses, the width of which can be regulated from the outside. Close to the slit, between it and the eyepiece, there is a total reflecting prism and the slit is illuminated by means of it and an opening at the side.

To adjust the instrument the eyepiece is first focussed on the slit and then a plane parallel plate of glass is fixed upon the prism table and arranged to reflect the rays from the telescope back into it. Thus an image is formed in the field of view above the slit itself, and by using the rack and pinion motion to make this image as sharp as possible the telescope is focussed for parallel light. To adjust it at right angles to the axis of rotation of the instrument the glass plate on the prism table is rotated through 180°, and the inclination of the telescope and levelling screws of the prism table altered in exactly the same way as in adjustment (*c*) of the last section, in order to bring the images of the slit both times into the same special position in the field. The adjustment of the refracting edge of the prism examined is made according to the first of the two methods (*d*) described in the last section.

To measure the angle A of the prism (fig. 102) the image of the slit is first obtained by allowing the light to fall perpendicularly on face AB. The prism table is then rotated so that the light falls perpendicularly on face AC. The angle rotated through is obviously equal to $\angle$MAN, and by subtracting this from π the value of A can be obtained.

To measure the index of refraction of the glass the slit is illuminated with sodium light and the rays allowed to fall on the face AC in

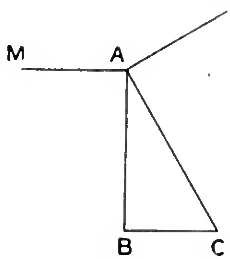


Fig. 102.

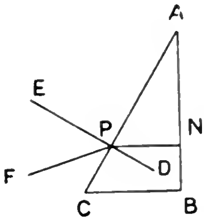


Fig. 103.

the direction FP, so that after refraction they fall on the face AB perpendicularly. They thus retrace their path after reflection and form an image of the slit in the correct position in the field. The prism table is then rotated, so that an image is formed by direct reflection on the face AC. The angle rotated through is obviously equal to $\angle$FPE. Since $\angle$DPN = $\angle$A, which has previously been measured, and μ = sin FPE/sin DPN, the index of refraction is therefore determined.

The advantage of the Abbe spectrometer is, that the moving parts are reduced to a minimum and there is less chance of the accuracy of the readings suffering from strain of the instrument. Also, as designed by Abbe, it had a micrometer screw for rotating the prism table, which enabled small differences of angle, e.g. dispersion in various parts of the spectrum, to be read with very great accuracy.

Total Reflection Methods for Determining Index of Refraction. To determine the index of refraction of solids and liquids by the methods already given it is necessary for the solids to be ground in the form of prisms and for the liquids to be contained in hollow glass prisms with plane parallel sides. It is also necessary for the solids and liquids to be transparent. In determining the index of refraction from the limiting angle of total reflection, however, a single drop of the liquid is sufficient; it is only necessary for the solid to have one polished face and the liquid or solid may be imperfectly transparent. The method can thus be applied to determining the index of refraction of milk or butter.

Let ABC be a glass prism, not necessarily equiangular, let the face BC be ground and the other faces be polished. Let us suppose that the image of a sodium flame is focussed on the ground face by a lens and that S is a point on this image. Then rays of sodium light diverge from S in all directions. Let μ be the index of refraction of the prism with reference to air, and let the face AB be covered with a liquid, the index of refraction of which with reference to air is μ_1.

Then, if a ray SP from S fall on AB at an angle of incidence ϕ given by $\sin\phi = \mu_1/\mu$, the sine of the angle of refraction is unity and the refracted ray emerges in the direction PA, grazing the face of the prism. If the angle of incidence is greater than this, e.g. SK, there is no refracted ray; the ray is totally reflected and the intensity of the reflected ray is equal to the intensity of the incident ray. If the angle of incidence is less than this, e.g. SH, there is a refracted ray and consequently the intensity of the reflected ray is less than the intensity of the incident ray. It is the same with the rays from the other luminous points on the surface BC. Consequently, if a telescope is focussed for parallel light and pointed in the direction RQ so as to receive the rays on their emergence from the prism, the field is divided into two parts by a line corresponding to the direction QR. One side of this line is illuminated by light which has been totally reflected and is consequently brighter than the other side. In applying the total reflection method the cross-wires are set on the line of separation where the intensity changes.

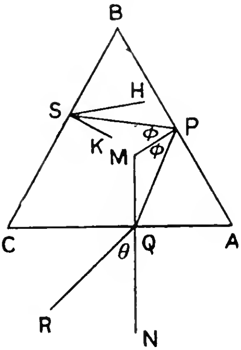


FIG. 104.

Let us suppose that instead of being luminous CB is covered with black paper to prevent the entrance of light, and that the rays from a sodium flame fall on the face BA in the direction BA at grazing incidence and at angles near grazing incidence. The ray at grazing incidence travels through the prism after refraction in the direction PQ. The other rays make smaller angles of refraction with the normal PM. Consequently, if the telescope be placed in the same position as before, the field on one side of the same line of separation is bright and on the other side is perfectly dark. It is here considerably easier to set the cross-wires on the line than in the other case where there was only a difference of intensity.

In fig. 104 $\angle A + \angle QMP = \pi$, also $\angle QMP + \angle MQP + \phi = \pi$, therefore $\angle A = \angle MQP + \phi$ or $\phi = A - \angle MQP$. Hence

$$\begin{aligned}\mu_1 &= \mu \sin\phi \\ &= \mu \sin(A - \angle MQP) \\ &= \mu \sin A \cos MQP - \mu \cos A \sin MQP.\end{aligned}$$

But, on applying the law of refraction at Q,

$$\sin\theta = \mu \sin MQP$$

and this gives on substituting,

$$\mu_1 = \sin A \sqrt{\mu^2 - \sin^2\theta} - \cos A \sin\theta \qquad . \quad . \quad (40)$$

If there is no liquid on the face AB, the critical angle of incidence is given by $\sin\phi = 1/\mu$, and this case is derived from the preceding one by putting $\mu_1 = 1$. Corresponding to (40) we obtain the equation

$$1 + \cos A \sin \theta = \sin A \sqrt{\mu^2 - \sin^2 \theta},$$

which gives $$\mu^2 = 1 + \left(\frac{\sin \theta + \cos A}{\sin A}\right)^2 \quad . \quad . \quad . \quad (41)$$

For putting the method into practice there are different types of apparatus, in which the scale reading gives the index of refraction direct, when the limiting angle is focussed on. The Abbe refractometer is an instrument of this type. Such instruments are usually calibrated with liquids of known refractive index. They cannot themselves be used to determine the A or μ of equation (40). They are, of course, very rapid in use. An ordinary spectrometer can, however, be adapted for use with the method, can be calibrated in the same way as any piece of special apparatus and will give quite as accurate results. It has the advantage that it can be used to determine A and μ, and then μ_1 can be calculated by equation (40) direct from first principles.

If an ordinary spectrometer is to be employed, the prism should have as high an index of refraction as possible, for its index of refraction must be higher than the indices of refraction to be measured. An extra dense flint equilateral prism such as is commonly used for producing spectra, with one face ground and an index of refraction for sodium light of about 1·6786, does very well. The first step is to determine μ and A, the angle between the two polished faces. This can be done by means of the methods described on p. 100. Then the prism table should be clamped and the prism fixed in position upon it, and if the index of refraction of a liquid is to be measured, a drop or two of it should be placed on AB (fig. 104) and a thin plate of glass pressed against that face of the prism. Capillary attraction keeps the plate in position and at the same time ensures that the drops are spread in a thin layer between the plate and the face of the prism. The limiting direction QR is then sought for with the telescope. The collimator is not used at all. If the light is to be incident internally, an image of a sodium flame should be formed with a lens on the ground face, and the beam forming the image should fall on that face so that its direction after refraction would be approximately SP. If the light is to be incident externally the beam of light should be directed on to the end of the plate, LM, as indicated in fig. 105. As a calculation will show, it cannot enter by the back of the plate, MN, and at the same time traverse the liquid in the required direction. If the end of the plate, LM, is polished, it must be plane and parallel to the refracting edge, A, of the prism.

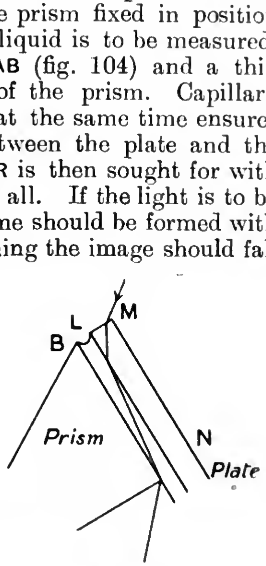


Fig. 105.

If the light is to be incident externally, it is much better to use instead of the plate another prism as similar to the first one as possible. Then the direction of the incident beam is always approximately parallel to the direction of emergence of the critical ray, so that to find the latter we have only to rotate the prism table in place of moving the light source and telescope independently.

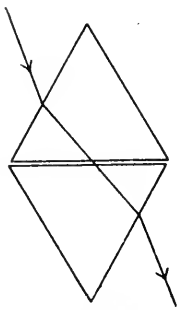


Fig. 106.

When the direction QR has been obtained, in order to find θ it is necessary to get the direction of the normal, QN. If the observations are made with a telescope fitted with a Gauss eyepiece, or if the spectrometer is an auto-collimating one, this can be done by means of the reflected image of the cross-wires or slit, as the case may be. If the instrument does not permit of this method, the position of the collimator on the divided circle should be read, before the experiment is started, by illuminating the slit, obtaining coincidence of the direct image of the latter with the cross-wires of the telescope, and subtracting 180° from the telescope reading. The collimator should then be kept fixed throughout the experiment. Then after QR is determined, the slit of the collimator should be illuminated again and the rays from the collimator allowed to fall on the face AC. If the position of the reflected image is then read by the telescope, the mean of it and the position of the collimator gives the direction of the normal. This follows since the angle of incidence is equal to the angle of reflection.

Some trouble may be experienced at first in getting the prism placed properly, but after the index of refraction has been obtained for one liquid, it can be done for others very rapidly.

If the index of refraction of a solid is to be obtained by this method, a polished face of it is placed against AB (fig. 104) with a drop of liquid between, of higher index of refraction than the index of refraction of the solid. The purpose of the liquid is to fill up the interstices, due to the two surfaces not being absolutely plane; the liquid commonly used is monobrom napthalene, the index of refraction of which for sodium light is 1·660.

Graphical Method of Determining the Index of Refraction of a Prism. There is a graphical method of determining the index of refraction of a prism, which does not require a spectrometer and is of great accuracy, one determination giving the second figure of the decimal and the mean of several determinations giving the third figure of the decimal.

Draw a straight line PQS (fig. 107) on a sheet of paper on a drawing-board, place the prism on the top of it and run a pencil along the two sides of the prism, AB and AC, so as to mark their positions on the paper. Then looking into the face AC of the prism draw TQ, so that it appears to be the continuation of PQ. Remove the prism and produce TQ to cut PS in Q. Cut off QS = QT and draw TR perpendicular to AC and SR perpendicular to AB. Join QR. Then the refractive index is given by

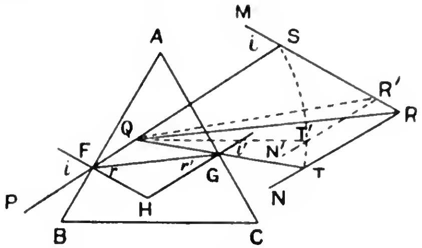


FIG. 107.

$$\mu = \frac{QR}{QS}.$$

For, if it is not, let it be given by

$$\mu = \frac{QR'}{QS},$$

and lay off the line QR' in the figure so as to meet SR. Through R' draw R'T' perpendicular to AC; with Q as centre and radius QS describe an arc to cut R'T' in T' and join QT'.

Let i be the angle of incidence of the ray PQ on the face AB of the prism and i' the angle of emergence of the ray QT from the face AC. Then $\angle MSQ = i$ and $\angle NTQ = i'$. Let the angles of refraction at the two faces of the prism, GFH and FGH, be denoted by r and r'. Then it may be shown that $A = r + r'$, and since SR and T'R' are perpendicular to AB and AC, $\angle SR'T' = A = r + r'$.

We have in $\triangle QSR'$

$$\frac{QR'}{QS} = \frac{\sin QSR'}{\sin SR'Q} = \frac{\sin i}{\sin SR'Q}.$$

But, by assumption,

$$\frac{QR'}{QS} = \mu.$$

Therefore

$$\mu = \frac{\sin i}{\sin SR'Q}$$

and $\angle SR'Q = r$. Consequently $\angle QR'T' = r'$.

In $\triangle QR'T'$

$$\frac{QR'}{QT'} = \frac{\sin QT'R'}{\sin QR'T'} = \frac{\sin QT'R'}{\sin r'}.$$

But

$$\frac{QR'}{QT'} = \mu.$$

Therefore

$$\mu = \frac{\sin QT'R'}{\sin r'}$$

and $\sin QT'R' = \sin i'$. Consequently $\angle QT'N' = i'$. But $\angle QTN = i'$, R'T' is parallel to RT, and QT' not parallel to QT. Hence $\angle QT'N'$

cannot equal i' and our assumption is wrong. The index of refraction must therefore be given by

$$\mu = \frac{QR}{QS}.$$

It will be noticed that it follows from the above proof that QR is parallel to FG, also the prism does not require to be in the minimum deviation position.

The Rainbow. Let SA be a ray incident on the surface of a transparent

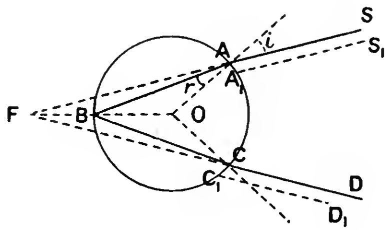


FIG. 108.

sphere in a plane through its centre. Then OA, the radius, is the normal at A. Let i and r be respectively the angles of incidence and refraction at A. After refraction the ray is reflected at the other side of the sphere at B, then passes to C and emerges into the air at C. Since OA = OB, the angle of incidence at B is r, consequently the angle of reflection OBC is r, $\angle OCB = r$, and the emergent ray CD makes the angle i with the normal at C. The deviation produced, namely $\angle SFD$

$$= 2\angle SFO = 2(\angle ABO - \angle FAB)$$
$$= 2(r - i + r) = 4r - 2i.$$

For this to be a minimum its differential coefficient with respect to i must $= 0$, i.e.

$$4\frac{dr}{di} - 2 = 0.$$

But $\sin i = \mu \sin r$, hence $\cos i = \mu \cos r \frac{dr}{di}$.

Eliminating $\frac{dr}{di}$ we obtain

$$\cos i = \tfrac{1}{2}\mu \cos r,$$

or

$$\cos^2 i = \tfrac{1}{4}\mu^2 \cos^2 r = \tfrac{1}{4}\mu^2 (1 - \sin^2 r) = \tfrac{1}{4}(\mu^2 - \sin^2 i)$$
$$= \tfrac{1}{4}(\mu^2 - 1 + \cos^2 i).$$

This can be rewritten

$$\cos^2 i - \tfrac{1}{4}\cos^2 i = \tfrac{1}{4}(\mu^2 - 1),$$

or

$$\cos i = \sqrt{\frac{\mu^2 - 1}{3}}.$$

If we substitute for μ its value for water for red light, namely, 1·329, we find that $i = 59{\cdot}6°$, $r = 40{\cdot}5°$, and the deviation is about 42·8°. If we substitute the value for violet light, namely, 1·343, $i = 58{\cdot}8°$, $r = 39{\cdot}6°$, and the deviation is 40·8°. The same result might have been obtained without the aid of the calculus by simply calculating r for different values of i and plotting the deviation as a function of i.

Now suppose that a second ray S_1A_1 is incident on the sphere parallel to the first. Its angle of incidence is slightly different from i, consequently in general its deviation is different from the deviation of SA, and after emerging from the sphere it is not parallel to CD. If, however, the deviation is plotted as a function of i, in the neighbourhood of the minimum it varies very slowly with i. Hence if SA is incident at the proper angle for minimum deviation, although the angle of incidence of S_1A_1 is slightly different, its deviation is almost the same as for SA, and after emerging from the sphere its path C_1D_1 is practically parallel to CD. We can have a great number of rays parallel to SA incident on the sphere between A and A_1 and after reflection they are still parallel and lie between CD and C_1D_1. We arrive therefore at the result, that if a parallel beam of light suffers minimum deviation in passing through the sphere, it emerges as a parallel beam. In all other cases it emerges as a convergent or divergent beam.

Now consider fig. 109. Let us suppose that the rays from the sun are

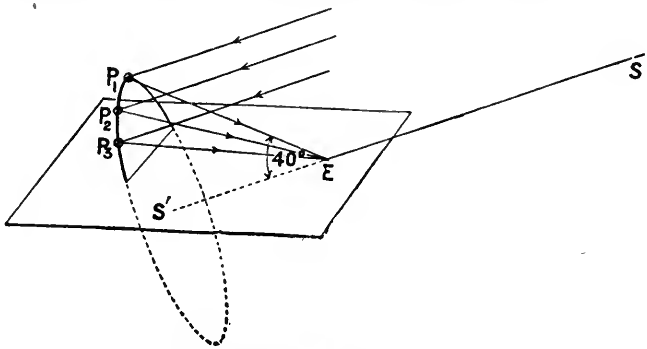


FIG. 109 (from Watson's "Physics").

incident in the direction SE, that an observer is situated at E, and that P_1, P_2, and P_3 are spherical rain-drops. If the rays from the sun are deviated by the rain-drops so as to arrive at the observer, he sees these directions bright, in which the rays suffer minimum deviation, i.e. which make an angle of about 40° with SE, 42·8° for red light and 40·8° for violet, with the other spectral colours coming in their order in between. Thus the rainbow is formed, and it is obvious that the red must be on the outside and the observer have his back to the sun. A rainbow formed in the above manner is called a primary rainbow.

If the rays are reflected twice inside the drop, as shown in fig. 110, for minimum deviation the red rays make an angle of about 51° and the violet rays an angle of about 54° with the incident light. Consequently a rainbow will be formed with the violet outside. This rainbow is called the secondary rainbow, and to see it the observer must also have his back to the sun. The primary and secondary rainbows are sometimes seen together, the secondary one being outside the primary one and fainter and broader than the latter.

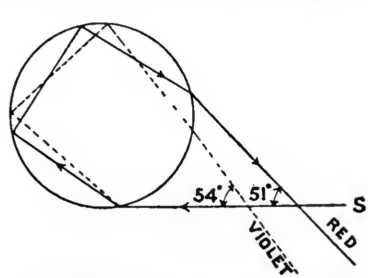


FIG. 110 (from Watson's "Physics").

Rainbows are also formed by rays which suffer three or four internal reflections, but such rainbows are very rarely seen. They are visible on the same side as the sun, but are faint in comparison with the scattered light from the latter, and are visible only when it is screened off by a bank of cloud.

The foregoing geometrical theory is principally due to Descartes. It does not completely explain the phenomenon. In addition to the bows already mentioned, there are alternations of brightness seen sometimes near the inner edge of the primary bow and more rarely at the outer edge of the secondary bow. These are called the supernumerary or spurious bows. Their explanation was first fully given by Sir George Airy, and depends on the difference of phase * of the different rays leaving the drop in directions near that of minimum deviation. The supernumerary bows are thus really a diffraction * effect due to the passage of the original plane wave through the drop.

The spray from a waterfall forms rainbows in exactly the same way as rain does.

EXAMPLES.

(1) Instead of taking the spectrometer telescope to a window and focussing it on an infinitely distant object, a student turns it towards a lamp in the room at a distance of 10 metres and gets the image of the latter sharp. The cross-wires are then approximately at a distance of 30 cms. from the object glass. How many millimetres must the object glass be racked in to make the focus exactly right?

(2) Show that when light passes through a thin prism the deviation does not vary with the angle of incidence, provided that the incidence is nearly perpendicular.

Show that in the same circumstances the deviation of the portion of the light that is reflected back from the second face of the prism differs from that of the light reflected back from the first face by a constant amount.

(3) A ray of light is refracted through a prism in a plane perpendicular to its edge. Prove that if the angle of incidence is constant, the deviation increases with the angle of the prism. Show that

$$\sin^{-1}\left(\frac{\sin i}{\mu}\right) + \sin^{-1}\left(\frac{1}{\mu}\right)$$

is the limiting angle of the prism such that the ray does not emerge when it meets the second face, i being the angle of incidence on the first face.

(4) Prove that for a prism of angle A

$$\frac{\sin \frac{1}{2}(A + \delta)}{\sin \frac{1}{2}A} = \mu \frac{\cos \frac{1}{2}(r - r')}{\cos \frac{1}{2}(i - i')},$$

where δ is the deviation, i and i' the angles of incidence and emergence, and r and r' the corresponding angles of refraction.

(5) A prism with a refracting angle of 60° is made of a glass, the refractive indices of which for Na and Li are respectively 1·5170 and 1·5140. A student measures the angle of minimum deviation for Na, and then, instead of setting the prism at minimum deviation for Li, measures the deviation of the Li line when the Na line is at minimum deviation. How much will his determination of the refractive index for Li light be out?

(6) A 60° flint-glass prism has an index of refraction for sodium light of the value 1·6499. The change per 1° C. rise of temperature is + 0·000003.

* These terms are explained in Chapters IX and X.

Given that the divided circle of a spectrometer reads to 10″ of arc, and that the temperature of the laboratory in which it stands varies in the course of the year from 50° to 70° F., should the values of the index of refraction as determined by the spectrometer show any appreciable variation?

(7) Make a graphical determination of the index of refraction of a glass prism by the method given on p. 110. The individual results should be accurate to the second decimal place.

(8) Work out the theory of the secondary rainbow in the same way as the primary is done on p. 112.

PART II.

PHYSICAL OPTICS.

CHAPTER VIII.

THE VELOCITY OF LIGHT.

Galileo made an attempt to determine the velocity of light by means of two observers furnished with lamps and situated a distance apart. The first observer uncovered his lamp and the second observer uncovered his as soon as possible after seeing the light from the first observer's lamp. The idea was, that the time which elapsed between the first observer's uncovering his own lamp and his seeing the second lamp would be equal to the time taken by the light to go from him to the second observer and back. The method failed owing to the enormous velocity of light, the time taken by it to travel the distance in question being very much less than the time necessary to uncover one lamp or to see another.

The velocity of light can be determined experimentally by four separate methods. These, taken in order of their discovery, are Römer's method, the aberration method, Fizeau's method, and Foucault's method.

Römer's method. Four of the moons of the planet Jupiter are large enough to be observed easily with a small telescope. Their periods of rotation about Jupiter vary from 42 hours to $16\frac{2}{3}$ days. Their orbits are in approximately the same plane as the orbit of Jupiter about the sun. They are, of course, dark bodies and are illuminated solely by the reflected light of the sun; consequently when they enter the shadow cast by Jupiter they are eclipsed or disappear. Now it is natural to assume that they rotate about Jupiter with uniform angular velocity. Consequently the interval of time that elapses between two successive eclipses of any one moon should always be the same.

In 1675, however, the Danish astronomer Römer observed a peculiar variation in the times of occurrence of the eclipses. When the earth was approaching Jupiter they occurred too close together, and, when the distance between the earth and Jupiter was increasing, they occurred too far apart. He explained the difference by means of the time taken by light to pass through space.

In fig. 111 let S represent the sun, and let the two circles be the orbits of the earth and Jupiter. E and J are the positions of the earth and Jupiter when they are in conjunction, i.e. nearest one another, and E′ and J′ their positions when in opposition, i.e. farthest

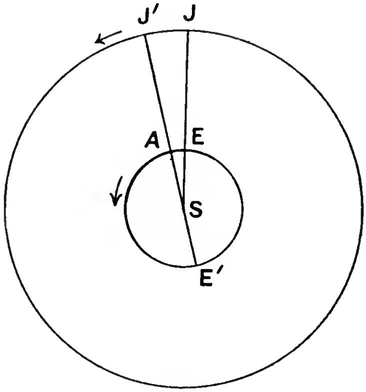


FIG. 111.

from one another. Jupiter takes 11·86 years to make a revolution round the sun, so that it moves only from J to J′ while the earth moves from E to E′.

We can regard the eclipses of any moon as a uniform series of time signals sent out by Jupiter. Owing to its distance from Jupiter the earth does not receive them until an appreciable time after they are sent out. As the earth moves from E to E′, its distance from Jupiter increases and the signals are received later and later; as the earth moves from opposition to conjunction again the distance decreases, the signals are received earlier, and when the earth and Jupiter are again in conjunction the lost time is made up. According to Römer, when the earth is in opposition, the signals have fallen 996 seconds behind the uniform rate and this is the time taken by the light to travel the distance AE′, i.e. the diameter of the earth's orbit.

The diameter of the earth's orbit was known only approximately in Römer's time. It is obtained from the solar parallax, the angle subtended by the earth's radius at the sun, and the determination of the latter is one of the most difficult problems in astronomy. According to the latest determinations it is 8·80″. Hence, using the more recent value of 1002 seconds instead of Römer's 996, we find for the velocity of light

$$\frac{360 \times 60 \times 60}{8{\cdot}8\pi} \; \frac{3963}{1002} = 1{\cdot}855\ 10^5 \text{ mls./sec.}$$
$$= 2{\cdot}98\ 10^{10} \text{ cms./sec.}$$

Since the velocity of light has been determined more accurately by other methods, the above equation is of importance as a means of calculating the solar parallax.

The Aberration Method. The apparent direction of the light from a star depends on the motion of the telescope. For example, in fig. 112 let OS be the true direction of a star, let the telescope be pointed in the direction OS, and let the telescope and observer be moving with velocity v in the direction OP. Then, when the light is passing down the telescope, the latter is moving sideways; consequently the path of the central ray *relative to the telescope* is shown by the dotted line QO_1, and, if the image of the star is to appear in the middle of the field on the cross-wires, the telescope must be pointed in the direction O_1Q. Let θ be the true direction of the star, let V be the velocity of light, and let t be the time taken by the light to travel down the telescope. Then $O_1Q = Vt$ approximately, since OO_1 is small, and $O_1O = vt$. Also in triangle QOO_1

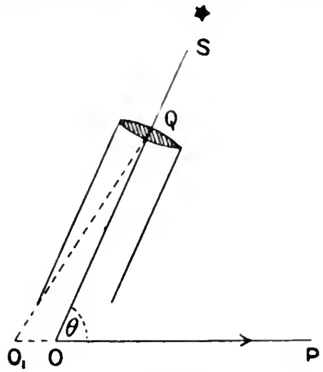


FIG. 112.

$$\frac{\sin O_1QO}{O_1O} = \frac{\sin QOO_1}{QO_1}.$$

This gives

$$\frac{\sin O_1QO}{vt} = \frac{\sin \theta}{Vt},$$

or, since sin O_1QO is very small,

$$\angle O_1QO = \frac{v}{V} \sin \theta.$$

Thus, owing to the motion of the telescope, the star is displaced in the direction of that motion in front of its true position by an angle equal to $(v \sin \theta)/V$.

The earth moves in its orbit about the sun with a velocity of about $18\frac{1}{2}$ mls./sec. If this value be substituted for v and sin θ be put equal to unity, $(v \sin \theta)/V$ takes the value 20″. Thus if the telescope is moving with the velocity of the earth, the stars receive an apparent angular displacement varying from 20″ to 0′ according to their position in the heavens.

This apparent displacement is known as aberration. It was discovered and measured in 1726 by Bradley, the Astronomer Royal, while looking for another effect. He gave the correct explanation, and calculated the velocity of light from his observations. The method suffers from the same disadvantage as Römer's, as the value for the velocity of the earth depends on the solar parallax. Bradley's discovery proved the correctness of Römer's views; until then they had been neglected.

The effect of aberration is to make the apparent position of each star execute an annual motion about its true position. If the star is in the ecliptic this motion is in a straight line ; if it is at the pole of the ecliptic, this motion is circular, and for all other positions it is in an ellipse.

The value for the aberration constant adopted at present as a result of observations is 20·47″.

It should be stated that while the simple theory stated here gives the correct value for the velocity of light there are difficulties in the way of its adoption. Cf. Chap. XXVI.

Fizeau's Method. The first terrestrial method of determining the velocity of light was carried out by Fizeau in 1849. His arrangement is shown diagrammatically in fig. 113. A beam of light from a source

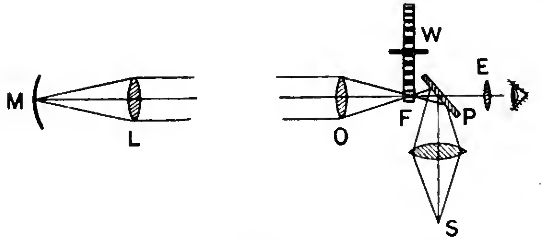


Fig. 113.

S passes through a converging lens system, is reflected by a glass plate P, and comes to a focus at F. It is then made parallel by the lens O, traverses a very great distance OL, falls on the lens L and is brought to a focus on the surface of the concave mirror M. The radius of curvature of this mirror is equal to ML, its distance from the lens ; the central ray of any pencil through the lens thus falls on the mirror normally and is reflected back the way it comes, even though inclined to the axis of the mirror. The lens and mirror L and M thus direct the beam back through the lens O to form a real image once more at F, and the observer looks at this real image through the eyepiece E and the glass plate P.*

W is a toothed wheel, and, as it rotates, its teeth pass one after another through the point F alternately stopping and letting through the light. If the wheel is moving slowly the eye sees a flickering image of S. If the images succeed each other faster than 15 or 20 a second the flickering ceases owing to the persistence of vision, and the image becomes steady. It is, of course, not as bright as it would

* The reflex lights sold for attaching to the back of bicycles consist of a lens and mirror arranged exactly as L and M. The lens is usually of red glass; the rays from the head lights of the overtaking motor car are reflected back in exactly the same way as in Fizeau's experiment, and the lens looks as if there were a faint independent light behind it. Reflex lights are, however, only of use in exceptional circumstances.

be if the wheel were away; the teeth in passing stop some of the light.

If the speed of the wheel is still increased, so that the time taken by the light to go from F to M and back is exactly equal to the time required for a tooth to move into the position formerly occupied by an open space, the light is intercepted by a tooth on its return and the image vanishes. If the speed of the wheel is now doubled, the light passes through the next space and the image is again visible; if the speed of the wheel is trebled, the light is intercepted by the next tooth and again vanishes. And so on; as the speed increases, the image alternately appears and vanishes.

In Fizeau's experiments the wheel had 720 teeth and the widths of the teeth and open spaces were equal. The distance between M and F was 8·6 km. A determination was made of the angular velocity of the wheel for which the image disappeared. Let it be ω radians/sec. for the nth disappearance, and let V be the velocity of light in kms./sec. Then

$$\frac{2 \times 8{\cdot}6}{V} = \frac{(2n - 1)}{2} \frac{2\pi}{720} \frac{1}{\omega},$$

i.e.

$$V = \omega \frac{720 \times 2 \times 8{\cdot}6}{(2n - 1)\pi},$$

whence V can be calculated.

Determinations of the velocity of light by Fizeau's method have been made by Cornu in 1874, Young and Forbes in 1881, and Perrotin in 1900. Cornu used a distance of 23 kilometres. Also, instead of observing the velocity for the disappearance of the image, he determined the velocities for which its brightness attained a certain value in diminishing to its minimum and afterwards in rising from its minimum. He determined the velocity by means of a chronograph which recorded seconds and tenths of a second, also every hundred revolutions of the wheel, and which was provided with a key under the control of the observer for recording any instant at which he desired to know the velocity. Cornu's result for the velocity of light in vacuo is $3{\cdot}004\ 10^{10}$ cms./sec.

Young and Forbes bevelled the teeth of the wheel in their apparatus so that the light stopped by the wheel was reflected to the side and lost. In Fizeau's apparatus it caused a general illumination of the field. They also silvered the plate P, leaving a small aperture for viewing the image, and altered the method of observing, employing two distances simultaneously.

Perrotin used Fizeau's apparatus as modified by Cornu and obtained for the velocity in vacuo the result $2{\cdot}9986\ 10^{10}$ cms./sec. His work was done at the Nice Observatory, and the distance FM (fig. 113) was 40 kilometres.

Foucault's Method. Fig. 114 shows the details of the method by which Foucault made a determination in 1862. It requires a much shorter distance than Fizeau's method. S is a rectangular aperture illuminated with sunlight, Q a plane parallel plate of glass,

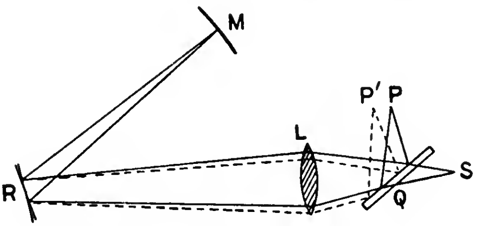


FIG. 114.

L a lens, R a plane mirror, which can be rotated about an axis perpendicular to the plane of the figure, and M a concave mirror. The lens L forms an image of S on M. The centre of curvature of M is at the centre of R, hence, no matter what the position of R is, if the light from it falls on M at all it falls on it normally and is thus reflected back along its path. If R is in the same position when the light reaches it again, the rays travel back to Q and are reflected to form an image at the side at P. If R is rotating rapidly, it has moved through an appreciable angle by the time the beam returns from M, and consequently the light is reflected in the direction of the dotted rays to form an image at P′.

In Foucault's experiment the distance RM was 20 metres and the displacement PP′ 0·7 mm. From the displacement the angle turned through by R was calculated and then, the angular velocity of R having been determined, the time taken by the light to go from R to M and back was known.

By placing a tube with water between R and M Foucault was able to show that the velocity of light is less in water than in air.

Foucault's method has been used and considerably improved by Michelson and by Newcomb. Michelson placed the lens between R and M and was able to increase the distance RM to 600 metres. His result (1882) was 2·9985 10^{10} cms./sec. and Newcomb's result 2·9986 10^{10} cms./sec., the same value as Perrotin has since obtained by the other method. Hence the latter value may be taken as correct. The two terrestrial methods are, of course, much more accurate than the astronomical methods.

EXAMPLES.

(1) In astronomy stellar distances are measured in "light years," i.e. the distance which light travels in a year, no star being as near the earth as three light years. How many miles are there is a light year?

(2) Given that the mean distances of the earth and Jupiter from the sun are respectively 93 and 483 million miles, find the maximum rate of increase of the distance of the earth from Jupiter. Hence find the greatest percent-

age diminution or increase in the periods of occultation of Jupiter's satellites. Note the distinction between the rate of increase of the distance of the earth from Jupiter and the relative velocity of the earth and Jupiter.

(3) In estimating the magnitude of the aberration on p. 121 we considered only the velocity of the telescope due to motion of the earth in its orbit round the sun. The aberration can be determined experimentally with the most accurate instruments to about $\frac{1}{20}$ second of arc. What appreciable effect, if any, will the diurnal rotation of the earth have upon it?

(4) Show that owing to astronomical aberration the apparent position of a star at the pole of the ecliptic traces out the hodograph of the earth's orbital motion. The pole of the ecliptic is the direction at right angles to the plane of the earth's orbital motion.

(5) If θ is the latitude of a star, i.e. its angular distance above or below the ecliptic, find in seconds of arc the semi-axes of the ellipse which its apparent position in the sky traces about its true position.

(6) In determining the velocity of light by Fizeau's method the distance is 10 km., the wheel has 720 teeth, and the angular velocities in radians per second for four successive disappearances of the image are 326, 457, 588, and 719. Find what disappearances these are and calculate the velocity of light from the data supplied.

(7) Make a reflex light and test it and consider whether they are worth using. A watch glass silvered on the back will do for the mirror. In the reflex lights sold the mirror is nickel-plated iron.

(8) Derive a formula for the velocity of light as obtained by the arrangement shown in fig. 114, in terms of the distances RM, RS, and PP′, and the angular velocity of the mirror, ω.

CHAPTER IX.

INTERFERENCE.

The Nature of Light. Hitherto we have made no assumption as to the nature of light. In the section dealing with Geometrical Optics we assumed only the rectilinear propagation and the laws of reflection and refraction. This was sufficient to explain everything with the exception of diffraction, which came in only on pp. 3, 84, and 114, and could, in general, be ignored. In order to explain the phenomena of interference it is necessary to make a further assumption.

We have seen that light travels with a finite velocity. Light rays carry energy with them. Now energy is propagated with a finite velocity in two different ways:—

(1) by the motion of the matter carrying the energy.

(2) by the transference of the energy alone, as in wave motion, the matter remaining stationary.

An example of the first way is the stream of bullets from a Maxim gun. There each bullet has an amount of kinetic energy given by $\frac{1}{2}mv^2$, where m is its mass and v its velocity, and if the bullet is stopped by a wall this energy reappears as heat in the wall. An example of the second way is the propagation of waves on the surface of water. When a wave passes, the water particles oscillate about their equilibrium positions, and after the wave is past they return to these positions. But the energy of the wave motion has moved on.

We explain things by means of analogy with other things, and consequently the attempt has been made to explain the propagation of light energy in both the above ways, by the emission theory and by the wave theory.

The wave theory of light was founded by Huygens, who published a treatise on light in 1690 after having stated the theory twelve years earlier. According to it, light is a wave motion propagated in the ether, the ether being a continuous medium in which matter exists and which fills all space. It is necessary to postulate the existence of the ether, for we cannot have wave motion without a medium, and there is no matter in the interstellar space through which the light comes from the sun.

In his book Huygens gave the following construction by which, if the position of the wave-front was known, its position could be found for any subsequent time: Consider all the points on the wave-front as centres of disturbance; then the envelope of all the secondary waves diverging from these centres gives the wave-front at subsequent times. For example, let CD (fig. 115) be a section of a wave-front which has diverged from the source and passed through the aperture AB. Then if it is desired to find the position of the wave-front after a time t, take in succession every point on the wave-front as centre and construct a sphere with radius equal to vt, where v is the velocity of light in the medium in question. All these spheres touch a surface of which EF is a section and this surface, the envelope of all the secondary waves diverging from the wave-front CD, is the wave-surface after time t.

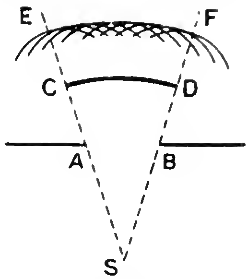


FIG. 115.

To apply Huygens' principle to the reflection of a plane wave let

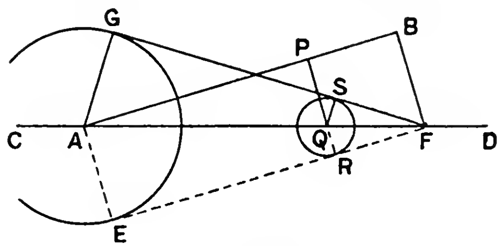


FIG. 116.

AB be the trace of a plane wave-front which is perpendicular to the plane of the paper, and let CD be the trace of a reflecting surface, also perpendicular to the plane of the paper. Draw BF perpendicular to AB to meet CD in F; draw AE parallel and equal to BF and join EF. Then EF gives a position which the wave-front would have occupied if there had been no mirror. Take P any point on the wave-front and draw PQR perpendicular to AB.

Let A and Q act as centres of secondary disturbances. With A as centre draw a circle of radius AE, and with Q as centre draw a circle of radius QR. Then, when the point B has reached F, the light from A and P has reached the circumferences of the two circles. Since FRE is tangent to both circles it is possible to draw a straight line FSG tangent to both circles. As P was any point on AB, FG touches the secondary waves derived from all points on AF, is hence the envelope of these waves and consequently the trace of the reflected wave-front. It is obvious from the figure that AB and GF are inclined equally to CD, and hence the angles of incidence and reflection are equal.

To apply Huygens' principle to the refraction of the same wave let v be the velocity in the upper medium and v' the velocity in the lower medium, let AB be the trace of the incident wave, let ED be the

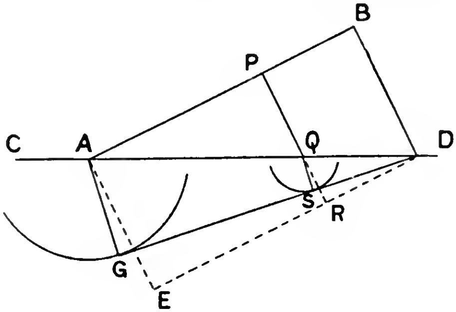


FIG. 117.

position it would have reached in time t, had there been no refraction, and let CD be the trace of the refracting surface. Then AE $= vt$. From P, any point on AB, draw PQR perpendicular to AB. With A as centre draw a circle of radius AG $= v't$, and with Q as centre draw a circle of radius QS $=$ QR v'/v. Then when the point B has reached D, the light from A and P has reached the circumferences of these two circles. Through D draw DG tangent to the larger circle; it may then be shown to touch the smaller circle and consequently to be the trace of the refracted wave-front.

AG gives the direction of the refracted ray and AE of the incident ray. Hence $\angle$ADG is equal to the angle of refraction and $\angle$ADE equal to the angle of incidence, and

$$\mu = \frac{\sin \angle ADE}{\sin \angle ADG} = \frac{AE/AD}{AG/AD} = \frac{AE}{AG} = \frac{v}{v'}.$$

This result is of fundamental importance. Put in words, it states that according to the wave theory the velocity of light in any transparent medium varies inversely as the index of refraction of the medium. Consequently the velocity in water and in glass is less than the velocity in air.

The wave theory of light as proposed by Huygens did not explain the rectilinear propagation of light, and hence was not accepted by Newton, who adopted the emission theory. According to it luminous bodies emit small material particles or corpuscles in all directions. These corpuscles fly in straight lines. Their mechanical impact on the retina causes the sensation of vision. When a corpuscle approaches within a very small distance of the surface of a refracting medium, according to its state at the time it is either repelled or attracted. Its velocity may be resolved into two components, one parallel to the surface and one normal to the surface. The parallel component in each case remains unaltered. If the corpuscle is repelled, the normal component of its velocity decreases to zero and then increases to its former value but in the opposite direction. The resultant velocity has thus the same value as before, its direction makes the same angle with the normal as the direction of the incident velocity does, and the reflection of light is explained in exactly the same way as the reflection of a perfectly elastic sphere by a wall.

If the corpuscle is attracted on approaching the refracting medium, it is bent towards the normal since the normal component of its velocity is increased. Also if v is the velocity in the first medium, v' the velocity in the second, i the angle of incidence, and r the angle of refraction, since the component velocity parallel to the surface remains unaltered $v \sin i = v' \sin r$ or

$$\frac{\sin i}{\sin r} = \frac{v'}{v},$$

i.e. is constant, for v and v' are both constants. Thus the correct law of refraction follows from the emission theory.

But if the parallel component of velocity is unaltered and the normal component increased on refraction, the resultant velocity is increased. The velocity of light should thus be greater in water and in glass than in air, exactly the opposite result to that obtained from the wave theory. Foucault's experiment on the velocity of light in water is thus an *experimentum crucis* between the two theories and it decides in favour of the wave theory. Owing to Newton's great authority the emission theory reigned supreme until the beginning of the nineteenth century, and it was the ability of the wave theory to give a satisfactory explanation of interference that then turned the scale against the emission theory.

We shall assume, then, that light is propagated by wave motion in the luminiferous ether. For the purposes of this chapter it will not be necessary to make any special assumption about the nature of this wave motion.

§ The rays are perpendicular to the wave-fronts. When an image

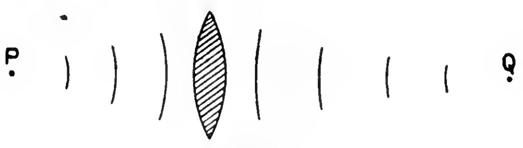


Fig. 118.

of a point P is formed by a lens, spherical waves diverge from P, and after passing through the lens become spherical waves converging towards the image Q. It thus follows from the wave theory that the optical length of the paths traversed by the different rays from P to Q is the same. This result might have been derived from the law of the extreme path, for since in the case under consideration the optical lengths of all the paths must be a minimum, it follows that they must be the same.

The formulæ for the position of the image formed by a lens or spherical mirror are sometimes derived from the property that all the paths have the same length, or from the property that one spherical wave-front is changed by the lens or mirror into another spherical wave-front with a constant difference of curvature. The latter method has recently become popular under the name of the "curvature" method. The method used in Chap. II of this book is however more fundamental, inasmuch as it uses only the laws of reflection and refraction.

§ To prove the formula for the focal length of a lens from the assumption that all paths from object to image have the same optical length let AB (fig. 119) be a lens, P an object, and Q the corresponding image. The lens is

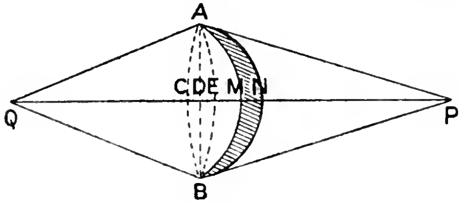


FIG. 119.

taken convex meniscus in form in order to have the diagram as clear as possible. AB cuts PQ at right angles in the point D. With P as centre and radius PA draw a circle to cut QP in C, and with Q as centre and radius QA draw a circle to cut PQ in E.

Then
$$QA + AP = QM + MN\mu + NP$$
$$= QE + EM + \mu MN + CP - CN$$

or
$$\mu MN + EM - CN = 0.$$

This may be rewritten
$$DE + CD = (\mu - 1)(DN - DM) \quad . \quad . \quad . \quad . \quad (1)$$

When two chords of a circle cut one another the rectangles contained by the segments are equal. Apply this theorem to the circle of which AEB forms part and let AB and QE be the two chords. The rectangle contained by the segments of AB is equal to AD^2. Since Q is the centre of the circle, the rectangle contained by the segments of the other chord is equal to $(2QE - DE)DE$, or approximately $2QE \, . \, DE$ since DE is small. Hence
$$2QE \, . \, DE = AD^2$$

or
$$DE = \frac{AD^2}{2QE} = -\frac{AD^2}{2v},$$

if v be written for QE, the image distance. Similarly writing u, r_1, and r_2 for the radii of the circles on which C, N, and M are situated, and using the same convention as to signs that is employed in Chap. II, we obtain
$$DM = -\frac{AD^2}{2r_2}, \; CD = \frac{AD^2}{2u} \text{ and } DN = -\frac{AD^2}{2r_1}.$$

On substituting these values in (1) we arrive at
$$-\frac{AD^2}{2v} + \frac{AD^2}{2u} = (\mu - 1)\left(-\frac{AD^2}{2r_1} + \frac{AD^2}{2r_2}\right),$$

which simplifies to
$$\frac{1}{v} - \frac{1}{u} = (\mu - 1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right).$$

Expression for a Light Wave. Consider the expression

$$\eta = b \cos \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) = b \cos 2\pi\left(\frac{t}{\tau} - \frac{x}{\lambda}\right).$$

By plotting it as a function of x for successive values of t it may be shown to represent an infinite train of progressive harmonic waves. The waves are said to be harmonic because they have the cosine form, the train is infinite as the expression gives real values throughout the whole range $-\infty < x < +\infty$, and the waves are said to be progressive, because, as t increases, the whole wave profile moves bodily forward in the direction of positive x. The wave-length, the distance between two successive crests at any instant, is given by λ; the period, that is the time taken by a complete wave to pass a fixed point, is given by τ; the velocity of the wave, v, is equal to λ/τ, and the amplitude of the wave is given by b. As mentioned above, in order to explain interference it is not necessary to make any special assumption as to what η represents, but for the sake of definiteness we may suppose that it denotes a displacement of an ether particle. Then the displacement is the same for all the ether particles on one wave-front, i.e. with the same value of x. As the wave passes the ether particles all describe simple harmonic motions at right angles to the direction of propagation.

As long as the wave remains in the same medium the square of the amplitude, b^2, is taken as a measure of its intensity. This may be taken as an assumption to be justified by experience, or it may be made plausible by analogy from other forms of wave-motion; the intensity of a light wave is proportional to the energy it brings with it, the kinetic energy of the vibrating ether particles is proportional to the square of their velocity, and their velocity is given by

$$\frac{d\eta}{dt} = -\frac{b2\pi}{\tau}\sin\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right).$$

Interference. If we have two systems of waves passing over the surface of water, each produces its own displacement of the surface independently of the other. This is known as the principle of the superposition of wave-motion. The resultant displacement of the surface is obtained by summing the two separate displacements. In the same way two light waves or beams cross without in any way interfering with one another. This was pointed out by Huygens. Different people can see different objects through the same aperture at the same time without any blurring due to the different trains of waves crossing one another.

Let us suppose we have two waves represented by

$$\eta = b \cos 2\pi\left(\frac{t}{\tau} - \frac{x}{\lambda}\right) \text{ and } \eta = b \cos 2\pi\left(\frac{t}{\tau} - \frac{x+\delta}{\lambda}\right)$$

going the same way. They have equal amplitude, wave-length, and period but a constant phase difference $2\pi\delta/\lambda$, that is, the crests of the one wave are always a constant distance δ ahead of the crests of the other. According to the principle of superposition the resultant effect is given by

$$\eta = b \cos 2\pi\left(\frac{t}{\tau} - \frac{x}{\lambda}\right) + b \cos 2\pi\left(\frac{t}{\tau} - \frac{x+\delta}{\lambda}\right)$$
$$= 2b \cos \frac{\pi\delta}{\lambda} \cos 2\pi\left(\frac{t}{\tau} - \frac{x + \delta/2}{\lambda}\right),$$

a wave of the same wave-length and period as the component waves but with an amplitude $2b \cos \pi\delta/\lambda$, which varies all the way from $2b$ through 0 to $-2b$ according to the magnitude of δ. Although the intensity of each component wave remains equal to b^2, the intensity of the resultant varies from $4b^2$ to 0. We thus have two light waves combining together to produce darkness. This phenomenon is called interference.

It is when $\pi\delta/\lambda = n\pi$, when $\delta = n\lambda$ or the crests of the one wave are an integral number of wave-lengths ahead of the crests of the other, that the intensity of the resultant wave is a maximum, and it is when $\pi\delta/\lambda = (n + \frac{1}{2})\pi$, i.e. when $\delta = (n + \frac{1}{2})\lambda$ or when the crests of the one wave coincide with the troughs of the other, that the resultant intensity is zero.

Let us suppose that we have two sources M and N (fig. 120) send-

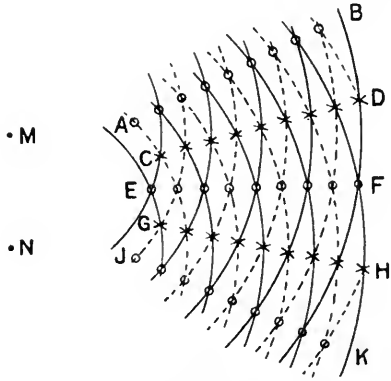


FIG. 120.

ing out waves of equal amplitude, equal velocity, and equal wave-length, and let the waves be represented by circles in the figure, the crests by full circles, and the troughs by dotted circles. The sources have the same phase, i.e. the two series of crests and troughs have the same distances from their respective centres. Then at the points marked by o's in the fig. we have crest superimposed on crest and trough superimposed on trough, so that the amplitude is twice that of each component wave. At the points marked by x's we have crest and trough superimposed on one another, the difference of phase is half a wave-length, the two waves neutralise one another, and the amplitude is zero. It must be noted that no energy is destroyed by the interference of the two waves; although the intensity is zero along the lines joining the x's, along the lines joining the o's it is four times as great as the intensity of a component wave.

It will be found by drawing the circles for any other time, i.e. by increasing the radii of all of them by the same amount, that GH and CD are always lines of zero intensity, and AB, EF, and JK always lines along which the intensity is increased four times.

§ Let two particles at M and N vibrate parallel to one another in the

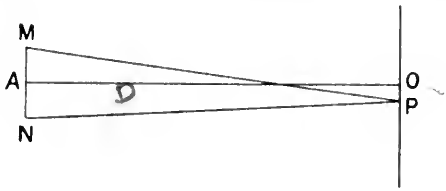


Fig. 121.

same phase and send out waves with the same amplitude and same period.

In order to consider the effect at a point P, join MN, and through A, the mid point of MN, draw AO perpendicular to MN. Draw PO perpendicular to AO. Let MN $= 2d$, AO $=$ D, and OP $= x$. Then

$$\mathrm{MP}^2 = \mathrm{D}^2 + (d + x)^2, \quad \text{and } \mathrm{NP}^2 = \mathrm{D}^2 + (d - x)^2,$$

therefore $\mathrm{MP}^2 - \mathrm{NP}^2 = (d + x)^2 - (d - x)^2 = 4dx$

and
$$\mathrm{MP} - \mathrm{NP} = \frac{4dx}{\mathrm{MP} + \mathrm{NP}}.$$

Let d and x be small in comparison with D. Then we have approximately MP $=$ NP $=$ D, and

$$\mathrm{MP} - \mathrm{NP} = \frac{2dx}{\mathrm{D}}.$$

If MP $-$ NP is equal to an integral number of wave-lengths, that is, if $2dx/\mathrm{D} = n\lambda$ or

$$x = \frac{n\mathrm{D}\lambda}{2d},$$

the two waves reinforce one another and we have a maximum of intensity at P, while if MP $-$ NP equals an odd number of half wave-lengths, that is, if

$$x = \frac{(n + \frac{1}{2})\,\mathrm{D}\lambda}{2d},$$

the two waves interfere and we have darkness at P.

So far we have considered only points in one plane. Let us suppose that there is a screen through OP perpendicular to the plane MPN and that it is required to find the distribution of the intensity on this screen. Draw a straight line PQ on the screen perpendicular to OP and let PQ $= y$. Then

$$MQ^2 = D^2 + (d + x)^2 + y^2,$$
$$NQ^2 = D^2 + (d - x)^2 + y^2,$$

and $$MQ^2 - NQ^2 = (d + x)^2 - (d - x)^2 = MP^2 - NP^2.$$

Thus, if PQ is short, MQ − NQ is approximately equal to MP − NP, and if there is a maximum of intensity at P there is also a maximum at Q.

Hence on the screen there is a number of parallel bright and dark bands, the distance between two consecutive bright or dark bands being

$$\frac{D\lambda}{2d}.$$

If instead of single particles at M and N we have pairs of corresponding particles arranged in two short lines perpendicular to the plane MON, if all the particles send out waves with the same amplitude and period, and if each pair of particles vibrates in the same direction and in the same phase, the systems of bands produced by every pair superimpose and the bands are consequently very much brighter.

Such bands are called interference bands.

Young's Experiment. Interference bands were first produced and explained in the above way by Young. He described his experimental

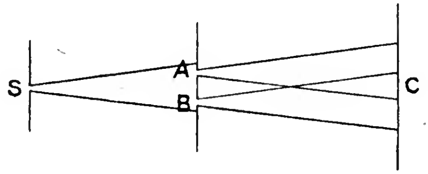


FIG. 122.

arrangement in his lectures which were published in 1807. A pencil of light was admitted through a slit S in a shutter. It then fell upon a screen in which there were two small pinholes, A and B, very close together. After passing through A and B the rays spread out and interference bands were formed on a screen at C where the two pencils overlapped.

Fresnel's Mirrors. In Young's experiment the two pencils which interfere pass through pinholes. Critics were thus not convinced that the bands were caused by interference of the two pencils; they were inclined to attribute them to a modification of the light produced by going through the holes, that is, to diffraction at the edge of the holes.

To remove this objection Fresnel devised two well-known arrangements for producing sources of light close together, namely, his mirrors and his biprism. These do not use any apertures or edges. Diffracted light was thus eliminated and the bands could be due to nothing else but interference.

In the first of these arrangements there are two plane mirrors inclined to each other at an angle of almost 180°, and light falls on them from a slit which is parallel to the line of intersection of their

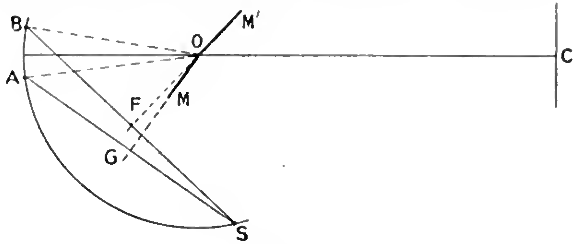


FIG. 123.

surfaces. Fig. 123 represents a plan of the arrangement; S is the slit and M and M′ are the mirrors. To find the positions of the virtual images of S formed by the mirrors produce M′O to F and OM to G; draw SF perpendicular to M′F and SG perpendicular to MG and lay off FB = FS and GA = GS. Then the images of S are situated at A and B; after reflection at the mirrors the light from S proceeds as if it came from A and B and interference bands are formed on a screen at C.

Let OS = a, let OC = b, and let the angle FOG be denoted by ω. It is obvious from the construction that A and B are on the circle with centre O and radius OS. Since FS and GS are respectively perpendicular to OF and OG, ∠FSG = ∠FOG = ω. ∠AOB stands on the same arc as ∠FSG and has its vertex at the centre, consequently it equals 2ω. The arc AB thus equals $2a\omega$, and, since ω is small, $2a\omega$ may also be taken as the distance between A and B measured along the chord. The straight line OC is drawn so as to bisect the angle AOB. In the notation of p. 133 the distance between the bands is given by $D\lambda/(2d)$; on substituting D = $a + b$ and $2d = 2a\omega$ we obtain in the present case

$$\frac{(a+b)\lambda}{2a\omega}.$$

Fresnel's Biprism. In the second arrangement Fresnel performed by refraction what he had done in the first by reflection. The light

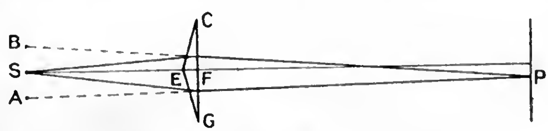


FIG. 124.

from the slit S falls on the biprism CEG, which is simply a prism with a very obtuse angle at E. It may be regarded as made up of prisms CEF and GFE placed base to base; hence the name. The edge E of the biprism divides the incident light into two portions, one which passes through the face EC and appears after refraction to come from B and another which passes through the face EG and appears after refraction to come from A. As the angles GCE and CGE are small and equal, A and B are situated close to S at equal distances from it. The effect of the biprism is thus to produce two virtual images of the slit which produce interference bands in the space to the right of CG, for example on a screen at P.

The measurement of the interference bands produced by a biprism is widely used as an exercise for students. A sodium flame is taken as the source of light, and the object of the experiment, as it is usually done, is to determine the wave-length of sodium light. The apparatus is an optical bench, usually of metal, with supports for the slit and biprism and for the vernier microscope used for measuring the distance between the bands. The bands are not produced on a screen at P but in the focal plane of this microscope. The slit and edge of the biprism are vertical, the tube of the microscope is parallel to the axis of the bench and has a horizontal motion at right angles to the latter, so that, when it moves, the cross-wires pass in succession across the different interference bands.

In setting up the apparatus the first requirement is to get the slit, biprism, and microscope in the same straight line. The slit is next made narrow, and then on looking into the microscope a vertical strip of light will be seen. This resolves into a number of equidistant parallel lines when the biprism is adjusted so that its edge is parallel to the slit, and when the width of the slit is further regulated. It is very important that the slit and edge of the biprism should be accurately parallel to one another.

On looking at the bands it will be noticed that some are much brighter than others, that there is a rhythmic variation in their brightness. This is due to diffraction, to the two halves of the biprism acting as rectangular openings. It is ignored by the theory of p. 133.

The formula for the wave-length is

$$\lambda = \frac{2df}{D},$$

where f is the distance between two successive bands, D is the distance from the virtual images to the focal plane of the microscope, and $2d$ is the distance between the two virtual images. Since the opposite faces of the biprism are almost parallel, the divergence of a thin pencil of rays passing through it is not appreciably altered. Consequently A and B are in the same plane as S and the distance D can be measured with a metre-stick. The best way to measure f is to read the position of ten or twelve, say ten, consecutive bands, subtract the reading for the first from the reading for the sixth, the reading for the second from the reading for the seventh, and so on, and divide the mean of the results so obtained by five.

There are two distinct methods of determining $2d$. In the first a convex lens is inserted between the biprism and the microscope and adjusted so as to form images of A and B in the focal plane of the latter. In order to do this it is usually necessary to move the microscope further away from the biprism, but the positions of the biprism and slit themselves must be kept unaltered. When the images are obtained, the distance between them is measured with the microscope. Let the result be c_1. The slit, biprism, and microscope are then kept fixed and the lens moved into the other position, in which it forms images of A and B in the focal plane of the microscope. Let the distance between the images in this case be c_2. Then $2d = \sqrt{c_1 c_2}$. The result follows since the magnification in the one position is the reciprocal of the magnification in the other.

In the second method the angles of the biprism FCE and FGE are measured with the spectrometer. Let the result be α. Then, since α is small and the light goes through the biprism almost normally, it may be shown that the deviation produced by each half of the biprism is $(\mu - 1)\alpha$. The angle between BE and AE in fig. 124 is thus $2(\mu - 1)\alpha$, and the distance between A and B consequently $2(\mu - 1)\alpha a$, where a is the distance of the biprism from the slit.

Fundamental Condition for Interference. In the three methods for producing interference just described—Young's experiment, Fresnel's mirrors, and Fresnel's biprism—it will be noticed that the two interfering sources are derived from the same source. Thus A and B are both illuminated by light from S in fig. 122, while A and B are merely images of S in figs. 123 and 124. This condition has been found indispensable for the successful production of interference bands. As an experimental fact it is impossible to get light from two independent sources to interfere.

Two different explanations have been given of this fact. According to the older explanation every source of light is subject to abrupt changes of phase very many times in a second, possibly owing to the molecule which is sending out the light vibration coming into collision with other molecules. Now in fig. 121 if the phase of M falls behind the phase of N, the central band of the system moves above O and the whole system of fringes is displaced. If the relative phase difference of M and N changes very many times in the second, the system of fringes is displaced many times in the second, too frequently for the eye to follow them ; the eye sees them superimposed and all trace of the individual bands is lost. But if M and N are derived from the same source, their phases change simultaneously, the system of bands remains stationary and is consequently always visible.

Schuster ("Optics," Chap. IV) has, however, pointed out that the harmonic waves, into which any actual wave can be analysed mathematically, never change phase. Another explanation is therefore necessary, and this he finds in the fact that we are never able to obtain perfectly monochromatic light. We have always to deal with a finite range of wave-lengths. If the two sources are independent the relative difference of phase of adjacent wave-lengths is never the same. Thus, for example, if the light used is that produced by the green line of thallium, the mean wave-length of which is $5{\cdot}3507\ 10^{-5}$ cms., we may have wave-lengths ranging from $5{\cdot}3537\ 10^{-5}$ to $5{\cdot}3477\ 10^{-5}$ cms., and the centre of the interference fringes produced by each component wave-length may lie at a different place on the screen. Consequently the different systems of fringes are blurred by superposition, give rise to a general illumination and no bands can be distinguished.

There is no contradiction between the two explanations since the wave-form contemplated by the older theory was the actual resultant wave and not one of its Fourier components. The two explanations are different ways of stating the same thing. It is, however, difficult to say what real existence the Fourier components have, unless the conditions of the problem in question are accurately specified. To appreciate Schuster's explanation one requires to be versed in Fourier analysis. The older explanation is less liable to give rise to misconception and is therefore to be preferred.

Lloyd's Mirror. Lloyd's mirror, first described in 1834, affords a method for producing interference bands which is both simpler and easier to work than Fresnel's mirrors or biprism. A mirror, MM′ (fig. 125), is placed so that the rays from a slit A fall on it at nearly grazing

FIG. 125.

incidence, and the reflected rays appear to diverge from a virtual image, B. The interference takes place between the rays from the slit and the rays from the virtual image, and the bands are examined with an eyepiece at P. The mirror may consist either of a piece of plate glass silvered on the front or a piece of unsilvered glass with the back blackened, so as to destroy the image formed by reflection in the second surface. The reflected rays cannot come below the plane of the surface of the mirror, so in general only one-half of the complete system of bands is visible. If, however, a thin plate of transparent material is placed in the path of the direct beam AP, the centre is displaced upwards and the complete system can be seen.

According to Dr. Lloyd the centre of the system does not lie in the plane of the mirror in the ordinary case but is displaced away from it by half the distance between two consecutive bands. The phase of the reflected light has thus been increased by π by the reflection.

There is one important difference between the two sources produced by Lloyd's mirror and those produced by Fresnel's mirrors or biprism. In the case of the latter the two images are similar, the right-hand side of the one corresponding to the right-hand side of the other and the left-hand side corresponding to the left-hand side. In Lloyd's mirror the right-hand side of the one source corresponds to the left-hand side of the other and vice versa. This property enables Lloyd's mirror to be used for the production of achromatic fringes.

The expression for the distance between two consecutive interference bands is $D\lambda/(2d)$, where D is the distance from the source to the screen and $2d$ is the distance between the two sources. It thus varies with the wave-length of the light employed. If white light is used, each wave-length forms its own system of bands. The central band of each system occupies the same position and hence the central band is white, but the other bands fall on different positions and blur one another. We thus have one white band surrounded by a few coloured bands and then general illumination.

If, however, by some means $2d$ could be made different and proportional to λ for the different colours, then the systems due to the different colours would superimpose exactly and the bands would be white, or in other words achromatic. This can be done by forming at A a spectrum of a narrow slit by means of a diffraction grating and using this spectrum instead of the slit at A. In a diffraction grating the deviation of the different colours is roughly proportional to the wave-length; hence if the spectrum be arranged at right angles to MM′ with the violet nearest the plane of the mirror, the violet in the image will also be nearest the mirror and by carefully adjusting the distance between the spectrum and its image d can be made proportional to λ.

§ Instead of using Fresnel's biprism two images of a slit can be produced for purposes of interference by a bi-plate or by Billet's half lenses. The figures illustrate how these act. In the first method two pencils of rays (fig. 126) from the slit S fall upon the bi-plate, one on

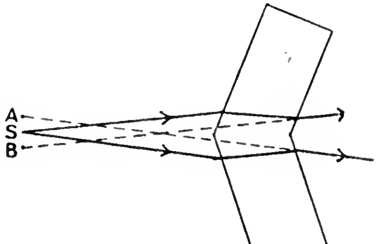


Fig. 126.

each half, and after passing through the latter they appear to come from the virtual images A and B. In the second method (fig. 127)

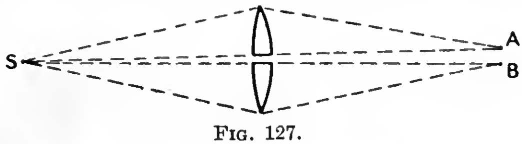


Fig. 127.

two real images A and B are formed of a slit S by means of a lens cut into two halves, the positions of which can be very accurately regulated relatively to one another.

Stokes' Treatment of Reflection. Let a ray AP of amplitude a be incident on the plane surface MK of a refracting medium. Then the amplitudes of the reflected and refracted rays PR and PC may be denoted by ar and ac, where r and c are each less than one.

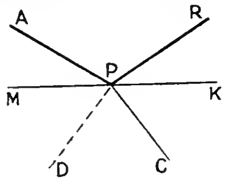


FIG. 128.

Let us suppose that the directions of the reflected and refracted rays are reversed. Then the reflected ray RP gives rise to a reflected ray along PA of amplitude ar^2 and a refracted ray PD of amplitude arc, while the refracted ray CP gives rise to a reflected ray along PD of amplitude acr' and a refracted ray along PA of amplitude acc'. But, when the reflected and refracted rays are reversed, they should combine to form the incident ray. Hence $a = ar^2 + acc'$ and $arc + acr' = 0$, which give $r' = -r$ and $1 = r^2 + cc'$.

Hence if we have two rays incident on the surface of a medium, one externally at a given angle of incidence and the other internally at the corresponding angle of refraction, the amplitude of the reflected ray has in the one case the same ratio to the amplitude of the incident ray as in the other but with a difference of sign.

The Plane Parallel Plate. A plane wave of amplitude 1 is incident in the direction AP_1 (fig. 129) on the surface of a plate of transparent

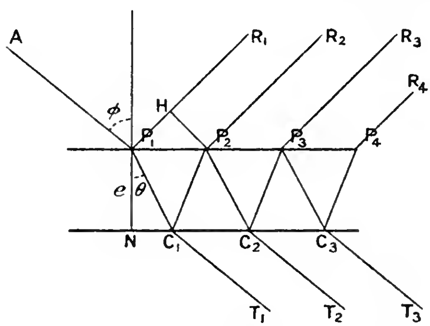


FIG. 129.

material with plane parallel sides and gives rise to a reflected wave in the direction P_1R_1 and a refracted wave in the direction P_1C_1. This refracted wave on arrival at the second surface of the plate gives rise to a reflected wave in the direction C_1P_2 and a transmitted wave in the direction C_1T_1. The reflected wave on its arrival at P_2 gives rise to a refracted wave in the direction P_2R_2 and a reflected wave in the direction P_2C_2, and so on; the original incident wave gives rise to a series of reflected waves and a series of transmitted waves of rapidly decreasing intensity. It is required to find expressions for the intensity of the resultant reflected and transmitted waves.

Let ϕ be the angle of incidence, θ the angle of refraction, e the thickness of the plate, let r, r', c, and c' have the same meanings as in the last section, and let μ be the index of refraction of the material of the plate. Let P_1N be perpendicular to the surfaces of the plate and draw P_2H at right angles to P_1R_1.

When the first and second reflected waves cross the plane HP_2 they are in different phases, for they were in the same phase at P_1 and they have traversed different optical distances in coming from P_1. Let their relative phase difference be denoted by d; then d denotes also the relative phase difference of any two successive reflected waves. We have

$$P_1C_1 = e/\cos\theta,$$

also $P_1H = P_1P_2 \cos HP_1P_2 = 2NC_1 \sin\phi = 2e\tan\theta\sin\phi = 2\mu e \sin^2\theta/\cos\theta$,

$$\text{hence}\quad d = \frac{2\pi}{\lambda}(\mu P_1C_1 + \mu C_1P_2 - P_1H) = \frac{2\pi}{\lambda}\left(\frac{2\mu e}{\cos\theta} - 2\mu e\frac{\sin^2\theta}{\cos\theta}\right)$$

$$= \frac{2\pi}{\lambda}\,2\mu e\cos\theta.$$

Let the incident wave be $\sin\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right)$ and let the resultant reflected wave be denoted by $R\sin\left\{\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) - \delta\right\}$, where the amplitude and phase, R and δ, are yet to be determined. Then the first reflected wave is $r\sin\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right)$, the second is $r'cc'\sin\left\{\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) - d\right\}$, the third is $r'^3cc'\sin\left\{\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) - 2d\right\}$, and

$$R\sin\left\{\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) - \delta\right\} = r\sin\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) + r'cc'\sin\left\{\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) - d\right\}$$
$$+ r'^3cc'\sin\left\{\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) - 2d\right\} \quad . \quad . \quad . \quad (2)$$

This holds for all values of $\left(t - \frac{x}{v}\right)$; hence, if it is expanded in terms of $\sin\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right)$ and $\cos\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right)$, the coefficients of each of these quantities is equal to zero, i.e.

$$R\cos\delta = r + r'cc'(\cos d + r'^2\cos 2d + r'^4\cos 3d + \ldots)$$
and
$$R\sin\delta = r'cc'(\sin d + r'^2\sin 2d + r'^4\sin 3d + \ldots).$$

The above two equations may be combined into

$$Re^{i\delta} = r + r'cc'(e^{id} + r'^2e^{i2d} + r'^4e^{i3d} \ldots)$$
$$= r + r'cc'\frac{e^{id}}{1 - r'^2e^{id}},$$

since the terms inside the bracket converge towards zero. On separating the real and imaginary parts the right-hand side of this equation becomes

$$r + \frac{r'cc'e^{id}(1 - r'^2e^{-id})}{(1 - r'^2e^{id})(1 - r'^2e^{-id})} = r + \frac{r'cc'(\cos d + i \sin d - r'^2)}{1 - 2r'^2 \cos d + r'^4}$$

$$= r + \frac{r'cc'(\cos d - r'^2)}{1 - 2r'^2 \cos d + r'^4} + i\frac{r'cc' \sin d}{1 - 2r'^2 \cos d + r'^4}.$$

Hence $R^2 = \left\{r + \frac{r'cc'(\cos d - r'^2)}{1 - 2r'^2 \cos d + r'^4}\right\}^2 + \left\{\frac{r'cc' \sin d}{1 - 2r'^2 \cos d + r'^4}\right\}^2,$

and this simplifies to

$$\frac{4r^2 \sin^2 \frac{d}{2}}{1 - 2r^2 \cos d + r^4} \text{ or } \frac{4r^2 \sin^2 \frac{d}{2}}{(1 - r^2)^2 + 4r^2 \sin^2 \frac{d}{2}}.$$

If T be the amplitude of the resultant transmitted wave, we have by the principle of energy

$$R^2 + T^2 = 1.$$

Hence $$T^2 = \frac{(1 - r^2)^2}{1 - 2r^2 \cos d + r^4}.$$

R^2 is 0 when $\sin^2 \frac{d}{2}$ is 0, i.e. when $\frac{d}{2} = \frac{2\pi\mu e \cos\theta}{\lambda} = n\pi$ or $2\mu e \cos \theta = n\lambda$, where n is an integer. Thus, if the path difference of two successive reflected waves is an integral number of wave-lengths, there is no reflected light. Again, R^2 can be written

$$\frac{4r^2}{\frac{(1 - r^2)^2}{\sin^2 \frac{d}{2}} + 4r^2}$$

and is hence a maximum when $\sin^2 \frac{d}{2} = 1$, or when the path difference is an odd number of half wave-lengths. The minimum value of R^2 is 0 and the maximum value is

$$\frac{4r^2}{(1 + r^2)^2}.$$

When R^2 is a minimum, T^2 has a maximum value equal to 1, and when R^2 is a maximum, T has a minimum and is equal to

$$\left(\frac{1 - r^2}{1 + r^2}\right)^2.$$

Approximate Theory. In the case of an unsilvered glass plate the second reflected wave is almost as bright as the first one, and the others are much fainter. Hence an approximation to the result can be obtained by considering these two alone and neglecting the others. The optical length of the path difference of the two waves is $2\mu e \cos \theta$; hence on first thoughts one would imagine that the two waves would reinforce one another and that there would be a maximum when $2\mu e \cos \theta = n\lambda$, where n is any integer. It was found on p. 140, however, that $r = -r'$, that external and internal reflection at the plate diminish the amplitude in the same ratio, but that in the one case this diminution is accompanied with a change of sign. This change of sign is equivalent to a relative phase difference of π or an alteration in the difference of the optical lengths of the paths by $\lambda/2$. Thus there is a maximum when $2\mu e \cos\theta = (n + \frac{1}{2})\lambda$, the same result as was obtained by consideration of all the reflected waves.

This result has important applications to the case of the colours of thin films, Newton's rings, and Haidinger's fringes.

Colours of Thin Films. If a thin film of oil is spread over the surface of water and viewed in reflected light, brilliant colours are often seen. They are due to the interference of light reflected at the upper and lower surfaces of the oil film, the latter acting as a plane parallel plate. The colours are especially brilliant if the film is on the surface of a muddy puddle on the road, for the transmitted light is then wholly absorbed and no light comes from the bottom of the puddle to dilute the purity of the interference effect. The same effect can also be obtained from the skin of a soap-bubble or from the thin paper-like pieces of glass that are sometimes the result of an amateur's glass blowing.

When we look at a small area on the surface of such a thin film, the rays diverging from it to the eye are almost parallel. The rays diverging from the front and back surfaces come originally from the same source, probably a white sky, and hence are in a condition to interfere. If $2\mu e \cos\theta = n\lambda$, the colour corresponding to that wavelength does not appear in the reflected light, and the film appears coloured. Owing to the rays received by the eye being not quite parallel, $\cos\theta$ is not the same for all of them. For one definite part of the spectrum to be blocked out, it is therefore necessary for n to be a small whole number and consequently for the film to be very thin.

Newton's Rings. If a lens, the surfaces of which have large radii of curvature, is placed on a glass plate and the point of contact viewed in white light, it is seen to be surrounded by coloured rings. These were first observed by Hooke in 1665, they were studied and their radii measured very carefully by Newton, and they were first explained satisfactorily by Young. They are due to interference between the light waves reflected at the upper and lower surfaces of the air film contained between the lens and plate.

In studying the formation of the rings Newton used telescope object glasses, the radii of curvature of which were many feet, and measured the radii of the rings formed by white light directly with the eye. In the experimental arrangement employed in laboratory courses nowadays spectacle lenses are used with a radius of curvature of perhaps 1 metre, the source of light is a sodium flame, and the diameters of the rings are measured with a vernier microscope. Either of two methods is employed. The air film is formed by the lower surface of the lens L and the upper surface of the plate P (figs. 130 and 131). In the first method the light from the flame passes

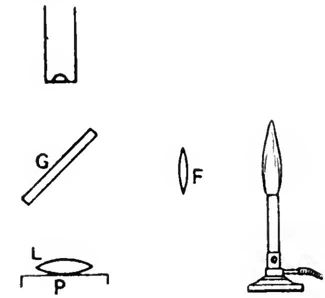


Fig. 130.

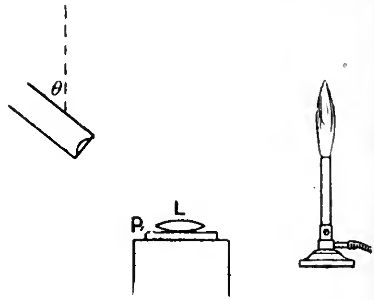


Fig. 131.

through the lens F, is reflected by the transparent glass plate G, and is brought to a focus on the film. It traverses the film vertically and is reflected by the surface of P, passes through G, and is then received by the microscope. The microscope is focussed on the air film. With this arrangement the rings are seen as a system of concentric circles. In the second method the microscope is inclined to the vertical at an angle θ and the light from the flame makes approximately the same angle with the normal to the upper surface of L. In this case, as they are seen obliquely, the rings appear as a system of ellipses.

The air film can be regarded as a plane parallel plate of slowly varying thickness. If R is the radius of curvature of the under surface of the lens, e the thickness of the air film at P, and CP $= \rho$ (fig. 132), then

$$\rho^2 = e(2\text{R} - e)$$

by considering the products of the segments of the horizontal chord

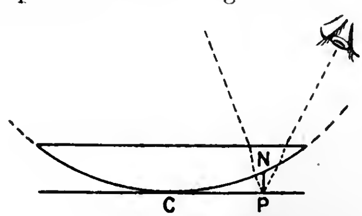


Fig. 132.

through N and of the vertical chord through C. Since e is small this equation may be written

$$e = \frac{\rho^2}{2\text{R}}.$$

Now consider the ray incident on the lens in the manner indicated by the dotted line. Let us suppose that the point P is so near C that the sides of the film may be considered parallel. Then, if $2\mu e \cos\theta = n\lambda$, the ray shown in the figure interferes with the ray reflected from the upper surface of the film, and to an eye situated above the film appears black at P. On substituting for e, the condition for interference becomes $2\mu\rho^2 \cos\theta/(2R) = n\lambda$ or

$$\rho^2 = \frac{n\lambda R}{\mu \cos\theta} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Hence on the ring with this radius we have blackness and this ring is one of a series, the radii of the others being obtained by giving n the other integral values. The squares of the first n radii are as the first n natural numbers. Similarly the radii of the bright rings, the loci of the points of greatest brightness, are given by

$$\rho^2 = \frac{(n + \frac{1}{2})\lambda R}{\mu \cos\theta} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Since the film is air, μ may be put equal to 1. Then, if ρ_n denote the radius of the n^{th} bright ring and ρ_{n+s} the radius of the $(n + s)^{\text{th}}$ bright ring,

$$\rho^2_{n+s} - \rho^2_n = \frac{(n + s + \frac{1}{2})\lambda R}{\cos\theta} - \frac{(n + \frac{1}{2})\lambda R}{\cos\theta} = \frac{s\lambda R}{\cos\theta},$$

whence

$$\lambda = \frac{(\rho^2_{n+s} - \rho^2_n)\cos\theta}{sR} \quad . \quad . \quad . \quad . \quad (5)$$

When the diameters of the rings have been measured by the method shown in fig. 130 or 131, the wave-length of sodium light can be calculated by the above formula. The case of fig. 130 is obtained from the general case by making $\cos\theta$ equal to 1. The advantage of using formula (5) instead of (4) lies in the fact, that if the lens or plate is deformed at the point of contact or if the contact is not perfect, a constant quantity is added to or subtracted from the theoretical value of the thickness, the whole system of rings is displaced, and for any one of them the square of the radius has not the value given by (4). Formula (5) is, however, unaltered, for it involves only the difference of the thicknesses at two points and this is still correctly given by the theoretical value.

The usual way of making the calculation is to measure the diameters of, say, the first 16 bands, take 8 for s and take 1, 2, 3, . . . 8 in succession for n, find the mean of the eight values of $\rho^2_{n+s} - \rho_n^2$ thus formed and substitute in the formula. R is found by the spherometer and θ is read from the microscope circle. The distinctness is usually improved by screening off part of the flame. This makes the area of the latter smaller and the light rays incident on the film more parallel. The rings can, of course, be observed by transmitted light but are then much fainter.

The central spot of the system is black. This is due to the half wave retardation introduced by the one reflection being an "external" one and the other an "internal" one. If, however, the lens is made of crown glass, the plate is made of flint glass, and the space between filled with oil of sassafras, which has an index of refraction intermediate in value between the indices of refraction of crown and flint glass, then in each case the light is reflected at the surface of an optically denser medium, the half wave retardation disappears, and the central spot of the system becomes bright. This experiment was first performed by Young.

If the light is white, the ring systems of the various constituent colours are superimposed. The result is a system of coloured rings, the colour at any point being not a pure spectral colour but including a considerable region of the spectrum. The order of succession of the colours from the centre outwards, seen by reflected light of course, was given by Newton as follows: (1) black, blue, white, yellow, red; (2) violet, blue, green, yellow, red; (3) purple, blue, green, yellow, red; (4) green, red; (5) greenish-blue, red; (6) greenish-blue, pale red; (7) greenish-blue, reddish-white. From the colours the thickness of the film can be obtained; thus if the incidence is perpendicular, the thicknesses corresponding to the reds at the ends of (1), (2), (3), (4), and (5), i.e. the reds of the first, second, third, fourth, and fifth orders, are $2{\cdot}5\ 10^{-5}$ cms., $5{\cdot}0\ 10^{-5}$ cms., $8{\cdot}0\ 10^{-5}$ cms., $10\ 10^{-5}$ cms., and $13\ 10^{-5}$ cms.

A good way of studying the colours of the rings in white light is by means of C. V. Boys' Rainbow Cup. This consists of a horizontal brass ring of about four inches diameter, which can be rotated rapidly about a vertical axis. A soap film is stretched across the ring, and when the latter rotates, owing to the centrifugal force called into play, the thickness of the film becomes very small at the centre and increases outwards from the centre to the edge. When the speed is properly adjusted a black spot can be seen at the centre surrounded by coloured rings, and the diameters of the surrounding rings can be altered by varying the speed of rotation.

Haidinger's Fringes. If instead of a thin film a plate of glass 3 or 4 mm. thick is taken, interference fringes formed by rays reflected from the front and back surfaces of the plate can be observed. Only it is necessary that the surfaces of the plate should be accurately plane and parallel, since the distance between the points where the interfering beams cut the surfaces is so much greater. Also the light must be monochromatic because for any given value of θ, owing to the large value of e, values of n can be found to satisfy every colour in the spectrum; thus if white light is used, the different coloured fringes combine to produce a general white illumination.

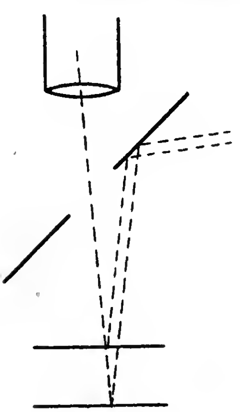


Fig. 133.

Such fringes were first observed by Haidinger (1849) and afterwards studied independently by Mascart and Lummer. They are best observed with a telescope focussed for infinity and pointing at right angles to the plate, the incident light being reflected by means of a mirror, as shown in fig. 133, so as to fall on the plate normally. If the plate is perfectly plane parallel the interference figures are then a system of concentric rings, the centre of which lies on the axis of the telescope. The shape of these figures can be used as a very accurate test of the parallelism of plates.

The accuracy of a plane surface can be tested more easily by placing it against a surface that is known to be optically plane, illuminating the air film thus formed with sodium light and examining the contour of the interference figures.

Applications of the Interference of Light. Besides their use in testing the parallelism of surfaces interference figures have also been employed in determining small changes in the index of refraction, in evaluating the wave-length of certain standard radiations in terms of the unit of length, and in studying the constitution of spectral lines. The application to the study of spectral lines will be left over to another chapter, but the other two applications will be described here.

The determination of the index of refraction of a substance by the method of minimum deviation or the method of total reflection reduces itself to the measurement of an angle. With a good spectrometer an angle can be read to 10 seconds and the index of refraction obtained to the fourth decimal place. The spectrometer is the most suitable instrument to use if the absolute value of the index of refraction of a solid or liquid is required. If, however, it is desired to measure a small change in the index of refraction of a solid or liquid, or if the substance under investigation is a gas, then other methods are more accurate. For suppose that a pencil of rays passes through the substance and that the actual length of the path in the substance is 10 cms., then if the optical length of this path changes by one wave-length, the index of refraction of the substance changes by $5\ 10^{-5}/10$, i.e. by 0·000005. Now it is easily possible to measure a change in the optical length of a path to one-fifth wave-length and hence by this method with a thickness of 10 cms. to obtain a change in the refractive index 100 times as accurately as with a spectrometer. Instruments embodying this principle are called interference refractometers or interferometers.

Fig. 134 explains the construction of Jamin's refractometer. A and B are two rectangular glass blocks of exactly equal thickness, made usually by cutting one block in halves. The rays diverging from a source S are made parallel by a convex lens and then fall on the surface of the block A, giving rise to a reflected beam C and a refracted beam. The latter, after reflection at the back of the block and refraction at the front, gives rise to the beam D. Both the beams C and D fall on the block B. If we consider only the portion of C that enters the block and is reflected at the back, and the portion of D that is reflected at the front, both beams superimpose after leaving B. They come from the same source and therefore are in a condition to interfere. The optical lengths of the two paths are not equal unless the blocks are parallel; hence if the block B is rotated in the manner

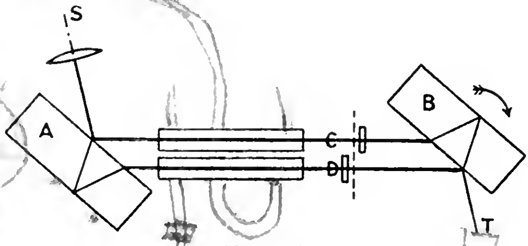


FIG. 134.

indicated in the figure, the relative phase difference constantly alters and interference bands cross the field of a telescope placed to receive the emergent rays at T. When the blocks approach the stage of absolute parallelism, the effect of inhomogeneity in the glass and inaccuracy in the surfaces begins to show and we obtain thick irregular figures. The back surfaces of the blocks A and B are silvered. The more nearly parallel the blocks are, the broader are the interference bands. Bands of this nature were first observed by Brewster.

If, for example, it is desired to measure the index of refraction of a gas, equal and similar tubes, closed at the ends with similar pieces of optically worked glass, are inserted one in the path of each beam, as shown in the figure. Initially the tubes are evacuated; then, by means of an arrangement not shown, the gas is allowed to slowly enter one of them. As it does so, the optical length of the path through that tube increases and interference bands move past the cross-wires of the telescope. The number that passes gives the increase in the optical length, and consequently, since the length of the tube is known, the index of refraction. The pressure at any stage can be read off from a manometer in connection with the tube.

Instead of counting the number of bands that pass, the same band may be kept on the cross-wires all the time by means of Jamin's compensator. This consists of two equally thick glass plates, the angle between which is kept fixed, and which can be rotated together about a horizontal axis indicated by a dotted line in fig. 134. The compensator is placed with one of its plates in the path of each beam. When the two beams make equal angles with the plates, the optical lengths of their paths through the plates are the same, but, when the plates are rotated, a relative phase-difference is introduced and this latter can be used to compensate the phase difference produced by the entering gas. It can be estimated from the angle through which the plates have been turned. The sensitiveness of the compensator can be varied by altering the initial angle between the plates.

In Rayleigh's refractometer (fig. 135) light from a slit S is made parallel by a lens, passes through the two tubes which are soldered together side by side, then through two slits and is finally brought to

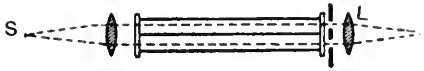


FIG. 135.

a focus by the lens L. The interference figures are formed in the focal plane of this lens and are examined by an eyepiece.

Michelson's Interferometer. Michelson's interferometer is capable of the same uses as Jamin's or Rayleigh's refractometers but has been used for other and more important purposes. Its principle is shown in fig. 136. The light from a source S separates at the back of the

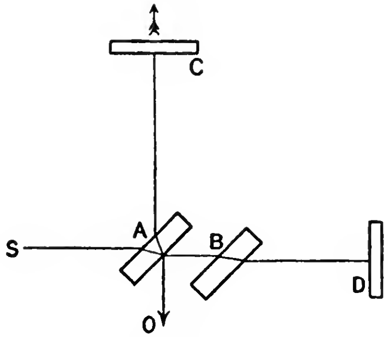


FIG. 136.

plate A into two beams, one of which goes to the plane mirror C and is then returned exactly on its path to O. The other passes through the glass plate B to the plane mirror D and is reflected back on its path to the back surface of A, where it is reflected to O. The two beams may be examined at O by a telescope. The back surface of A has a thin layer of silver deposited on it, so as to reflect about as much light as it transmits. The glass plate B is introduced to make the paths of the two beams symmetrical, as otherwise the beam to C would twice pass through a glass plate while the path of the other beam would lie wholly in air. The two plates A and B are placed parallel to one another and are worked originally in a single piece, which is afterwards cut in two. Both C and D are silvered on their front faces.

Fig. 137 shows one construction of the instrument. The mirror D is held by springs against three adjusting screws. The mirror C is mounted on a metal slide which can be moved along ways by the screw shown in the figure. The direction of the motion is shown by the arrow in fig. 136. It is very essential that parallelism of this mirror should be maintained when it is moved from one position to another.

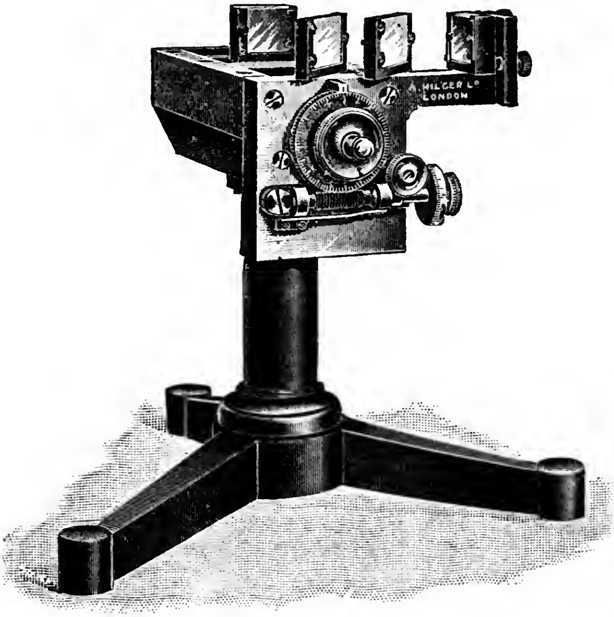


FIG. 137.

To adjust the instrument a small object, for example a pin, is held up between the source and the plate A. The observer at O sees two images of this pin, one formed by each beam. The fringes appear when these images are made to coincide. When white light is used, the fringes appear only when the paths AC and AD are equal.

The standard of length at the base of all scientific measurement is the standard metre, which is defined as the distance between two lines ruled on a metal bar made of an alloy of platinum and iridium. This bar is kept at the International Bureau of Weights and Measures at Sèvres. This standard is, of course, an arbitrary one; it is conceivably possible that in the course of time it and the copies made of it, which are in the same metal, might change by an extremely small fraction of a millimetre. The wave-lengths of spectral lines are invariable natural units. So if the standard metre were evaluated in terms of a certain definite wave-length, it could be checked at future times. This was done by Michelson in 1892. The wave-lengths chosen were those of the red, green, and blue lines of cadmium, which are extremely homogeneous; broad lines or lines accompanied by satellites will, of course, not produce satisfactory interference fringes. The details of the method would take too long to describe, but it involved measuring the distance between fixed plane mirrors on certain "etalons". The etalon was placed on the interferometer in place of the mirror D and the movable mirror C gradually moved from coincidence with the one mirror of the etalon to coincidence with its other mirror. The number of fringes that passed during the operation was counted, and the result gave the distance between the two mirrors of the etalon in wave-lengths.

The final result of the investigation was, that the number of light waves in a standard metre was found to be for the red radiation of cadmium 1,553,163·5, for the green 1,966,249·7, for the blue 2,083,372·1—all in air at 15° C. and at normal atmospheric pressure.

Lippmann's Colour Photography. If on the light wave

$$b \cos \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right)$$

progressing in the $+x$ direction we superimpose an equal wave

$$b \cos \frac{2\pi}{\tau}\left(t + \frac{x}{v}\right)$$

progressing in the $-x$ direction, the resultant disturbance is given by

$$b \cos \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) + b \cos \frac{2\pi}{\tau}\left(t \quad \frac{x}{v}\right)$$

which reduces to

$$2b \cos \frac{2\pi t}{\tau} \cos \frac{2\pi x}{\lambda},$$

where λ is written for $v\tau$. If this expression is plotted as a function of x for different values of t, it is found that it oscillates from the form

$$2b \cos \frac{2\pi x}{\lambda}$$

through a straight line to the form

$$-2b \cos \frac{2\pi x}{\lambda}$$

and then back again through a straight line to the first form. The displacement is always zero at the points given by $\cos 2\pi x/\lambda = 0$, i.e. by $x = (2n + 1)\lambda/4$, where n is any integer, and it has its maximum value at the points given by $\cos 2\pi x/\lambda = \pm 1$, i.e. by $x = n\lambda/2$. The former points are called nodes and the latter are called loops or antinodes. The expression as a whole is said to represent stationary waves because the wave does not progress; the crests keep on appearing and disappearing at the same points.

If light is reflected perpendicularly by a perfectly reflecting surface, stationary waves are formed, and if a phase change of half a wave-length occurs at the surface, there is a node at the surface and also at distances of $\lambda/2$, λ, $3\lambda/2$, etc., from it. Between the nodes there are loops. In 1891 Lippmann introduced a system of colour photography depending on this fact. A photographic plate was placed in a camera with the glass side facing the lens and the other side of the sensitive film was backed with a reflecting layer of mercury. The light entered the film from the glass, was reflected by the mercury and passed back through the film to the glass again. Inside the film it formed stationary waves. At the nodes there was no displacement and consequently no photographic action; at the loops there was a maximum of photographic action. When the film was developed and fixed, the silver in it was reduced in laminæ in the antinodal planes and not uniformly through the whole thickness of the film. Thus after development the film consisted of a great number of equidistant reflecting planes all parallel to the mercury surface and distant $\lambda/2$ from one another. The thickness of silver in each reflecting plane was so small that they were transparent.

Suppose now that white light is incident on this system of planes, then each plane gives rise to its own reflected wave. If we consider the different constituent colours of the white light separately, the reflected waves reinforce one another in the case of that colour for which half the wave-length equals the distance between the planes. For all other colours the reflected waves interfere and destroy one another. Thus when the plate is viewed normally in white light, each part of it appears in the colour of the radiation to which it has been previously exposed, and the whole plate gives an image of the original object in its natural colours.

Microscopic sections of the films have been prepared and the reflecting laminæ rendered visible to the eye. Lippmann's process is a very difficult one to carry out and has consequently been little used. It requires special plates with a grain finer than that of the commercial plates.

EXAMPLES.

(1) Supply the missing part of the proof on p. 128. Show rigorously that if DG (fig. 117) touches the large circle at G it also touches the small circle at S.

(2) Show that according to the corpuscular theory of light, when a corpuscle is incident upon an optically dense medium, the increase in the normal component of its velocity is $v \sin(i - r)/\sin r$, where v is the resultant velocity in the initial medium and i and r are the angles of incidence and refraction.

(3) Show that when a light corpuscle is refracted its gain in kinetic energy is independent of the angle of incidence. Hence the refraction of a corpuscle from air to glass can be regarded as due to a force of attraction, which acts in the direction of the normal, has a constant value F within a thin layer immediately above the surface of the glass, and is zero outside this layer.

Show that by analogy with a projectile under gravity the corpuscle passes in a parabolic arc from its straight flight in air to its straight flight in glass. Also show that in the case of total reflection at a glass-air surface the totally reflected corpuscle enters the air, describes a portion of a parabola in it and returns again to the glass.

(4) Derive the formula for the spherical mirror in the same way as the formula for the lens is derived on p. 130.

(5) Derive the formula for refraction at a spherical surface in the same way as the formula for the lens is derived on p. 130.

(6) Lord Rayleigh has given the following simple method of seeing Young's interference bands. Two small pieces of silvered glass are taken and a fine line ruled on one of them and two fine lines ruled as close together as possible on the other. The ruling removes the silver and the lines act as slits. The pieces of glass are then mounted in a short tube, which is held with the double slit close up to the eye. When the slits are parallel the bands are seen.

Show how to calculate their angular separation. (Instead of silvered glass surfaces the lines may be ruled on photographic plates which have been exposed and developed until they are black all over, or for the slits dense negatives may be made from black lines drawn on a white card.)

(7) Interference bands are produced by a Fresnel's biprism in the focal plane of a reading microscope. The focal plane is 100 cms. distant from the slit. A lens is inserted between the biprism and microscope and gives two images of the slit in two positions. In the one case the two images of the slit are 4·05 mm. and in the other case 2·90 mm. apart. If sodium light is used, find the distance between the interference bands.

(8) A plane wave of light is incident on the upper surface of a plane parallel glass plate at such an angle that the amplitude of the reflected wave is ·4 of the amplitude of the incident wave. Find the amplitude of the wave transmitted through the plate. Disregard all the internal reflections except the first.

(9) A plane wave of amplitude unity is incident on a plane parallel plate at such an angle that r, the amplitude of the first reflected wave, is 0·30. Derive an expression for the amplitude of the resultant transmitted wave and calculate the numerical values between which it varies, neglecting the variation of r with the angle of incidence.

What would these values be if we took only the first two transmitted waves into consideration?

(10) If the fraction of light reflected at the first surface of a parallel plate be r (there being no regular interference), that transmitted by the first surface, reflected by the second and again transmitted by the first is $r(1 - r)^2$. That reflected three times and transmitted twice is $r^3(1 - r)^2$, etc. Hence the whole reflected light is $\mathrm{R} = r + (1 - r)^2(r + r^3 + r^5 \ldots) = 2r/(1 + r)$.

(11) Newton's rings are formed between a plane surface of glass and a lens. The diameter of the fifth black ring is 9 mm. when sodium light is used, and the light passes through the air film at an angle of 30° to the normal. Find the radius of the glass lens.

(12) A convex spherical lens, the radius of curvature of the under surface of which is 20 cms., rests on a concave cylindrical lens, the radius of curvature of the upper surface of which is 40 cms., contact taking place at the lowest point. The combination is used to produce Newton's rings in the arrangement shown in fig. 130. What is the shape of the rings in this case?

(13) A soap film illuminated by white light gradually becomes thinner as the liquid drains away. It is placed in front of the slit of a direct vision spectroscope, which is held in a stand so that the slit is horizontal. Describe and explain the phenomena which are observed.

(14) Show that the phase difference between the two rays in a Jamin's interferometer is given by $\frac{4\pi\mu e}{\lambda}(\cos\theta_1 - \cos\theta_2)$, where θ_1 and θ_2 are the angles of refraction in the two blocks of glass and e is their thickness.

Two plates of glass of 1 cm. thickness, $\mu_D = 1·526$ and coefficient of linear expansion 0·000008, are placed one in the path of each ray. The temperature of one of the plates of glass being raised 1° C., 14 interference bands cross the field of view. What has μ_D become?

CHAPTER X.

DIFFRACTION.

HUYGENS' construction (p. 127) explained satisfactorily the reflection and refraction of a plane wave. In the form in which he gave it, it was unable to explain the rectilinear propagation of light. It was also unable to explain why the wave was not propagated backwards. For example, in fig. 138, if AB is a wave front diverging through a hole in a screen from a source S, and if with different points on AB as centres circles are drawn to represent secondary waves diverging from these points, these secondary waves travel round behind the screen at G and H. But there is darkness behind the screen at G and H. We have thus an apparent contradiction between theory and experiment. Also, the secondary waves touch in the surface EF as well as in the surface CD. If every point in the wave front is to be regarded as a secondary source, why is a wave not propagated backwards as well as forwards?

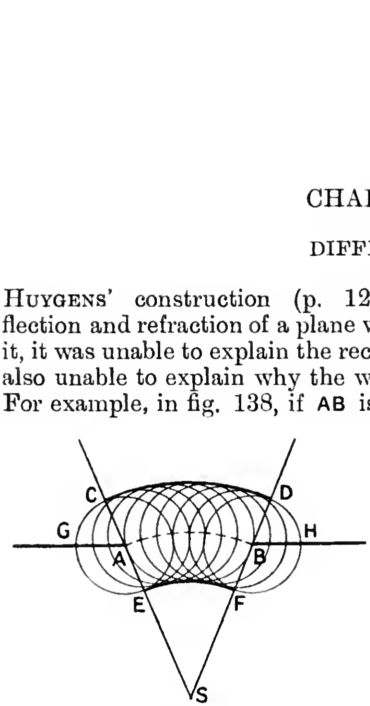


FIG. 138.

This difficulty about the light waves not bending round corners was one that the wave theory took a long time to get over. It was all the more pressing because in the analogous case of sound, which was known to be propagated by waves, the shadows were never sharp and the waves did bend round corners. The difficulty was explained away by Fresnel.

Fresnel's Explanation of the Rectilinear Propagation of Light. Let AB be a plane through which a plane wave is passing from left to right, and let P be a point at which it is desired to ascertain the effect of the wave. It is assumed that the light is monochromatic and that the wave consists of an infinite train of sine waves.

Draw PO perpendicular to AB and let PO = p. O is said to be the pole of P. Then with centre P and radii $p, p + \frac{1}{2}\lambda, p + \lambda, p + \frac{3}{2}\lambda$, etc., describe spheres. These spheres cut AB in circles, of which two are drawn—those cut by the spheres of radii $p + (n - 1)\lambda/2$ and $p + n\lambda/2$. A straight line is drawn through O to cut the circles in K_{n-1} and K_n. Then the area of the zone comprised between the two circles is

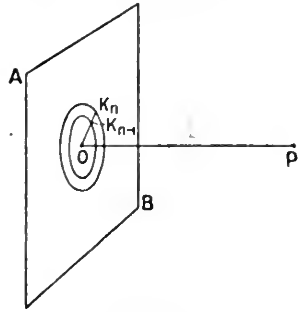


FIG. 139.

$$
\begin{aligned}
&\pi(OK^2_n - OK^2_{n-1}) \\
&= \pi(\{PK^2_n - p^2\} - \{PK^2_{n-1} - p^2\}) \\
&= \pi(PK^2_n - PK^2_{n-1}) \\
&= \pi\left(\left\{p + \frac{n\lambda}{2}\right\}^2 - \left\{p + \frac{(n-1)\lambda}{2}\right\}^2\right) \\
&= \pi p\lambda,
\end{aligned}
$$

if we neglect the λ^2 terms in comparison with the others. The area of the zone intercepted between any two successive spheres is thus always the same.

Let us regard every point on the plane as being in a state of vibration and sending out waves. Then the light vibration at P is due to the superposition of these waves. Since the incident wave is a plane one, all the points on AB are in the same phase. Their distances from P are different however, and consequently the secondary waves which they send out arrive at P in different phase. The distance of P from points in the first zone lies between p and $p + \frac{\lambda}{2}$, from points in the second zone between $p + \frac{\lambda}{2}$ and $p + \lambda$, and so on; consequently if the resultant amplitude due to points in the first zone is positive, that due to points in the second zone is negative, that due to the third zone is positive, and so on, and S, the resultant effect at P, can be written in the form of a series

$$S = m_1 - m_2 + m_3 - m_4 \ . \ . \ . + (-1)^{n+1} m_n,$$

where the successive terms represent the effects of successive zones and m_1, m_2, m_3, etc., are all of the same sign.

It has been shown above that the zones have equal areas, at least for small values of n; for large values the area increases very slowly with n. The amplitude of a light wave varies inversely as the distance from the source, and the distance of the zones increases with n. This increase in distance more than compensates for the slight increase in area. Hence each term of the series is slightly smaller than the one before it.

To sum the series we use the following method which is due to Schuster. Assume that the last term of the series is odd. The terms can be grouped in two different ways,

$$S = \frac{m_1}{2} + \left(\frac{m_1}{2} - m_2 + \frac{m_3}{2}\right) + \left(\frac{m_3}{2} - m_4 + \frac{m_5}{2}\right) + \ldots$$

and $$S = m_1 - \frac{m_2}{2} - \left(\frac{m_2}{2} - m_3 + \frac{m_4}{2}\right) - \left(\frac{m_4}{2} - m_5 + \frac{m_6}{2}\right) \ldots$$

Then, if each term is greater than the arithmetical mean of the preceding and following ones, the brackets are all negative, and the above two equations can be written

$$S < \frac{m_1}{2} + \frac{m_n}{2},$$

and $$S > m_1 - \frac{m_2}{2} - \frac{m_{n-1}}{2} + m_n.$$

But since the terms of the series diminish very slowly in size, m_1 can be written for m_2 and m_n for m_{n-1}. The limits between which S is enclosed become equal and consequently

$$S = \frac{m_1}{2} + \frac{m_n}{2}.$$

The cases when n is even and when each term is less than the arithmetical mean of the preceding and following terms can be treated in the same way. We thus arrive at the general result, that the effect of the whole wave at P is equal to half the effect of the first and last zones.

It is necessary for the validity of this result that the terms in the series should diminish regularly and very slowly.

In fig. 140 let the square drawn with the full lines be an aperture

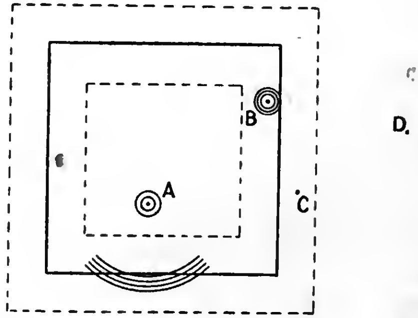


FIG. 140.

in a screen and let a plane wave fall on the aperture in a direction at right angles to the plane of the screen. Let A be the pole of a point at which it is desired to find the effect of the light wave. A is a considerable distance from the edge of the aperture, so that, if zones are drawn round A, by the time they intersect the edge they have a large diameter and are very thin. Hence the diminution in effective area takes place slowly, the terms in the series diminish slowly and the above result can be applied. The last zone may be drawn entirely off the aperture. Thus the effect of the whole wave reduces to half the effect of the first zone at A. Similarly, if the pole is at D, the effect reduces to half the effect of the first zone round D, which of course is zero, and we get no light at all.

But if the pole is at B inside the aperture and near its edge the effective area of the zones diminishes rapidly and the above theory does not apply. Similarly, if the pole is at C outside the aperture near the edge, the effective area of the first zones that intersect the edge increases too rapidly for the theory to apply.

As a result, therefore, we can state that there is uniform illumination at all points whose poles lie inside the inner dotted square, and that there is perfect darkness at all points whose poles lie outside the outer dotted square. As to the points whose poles lie between the two dotted squares, our theory as yet gives no information. Owing to the smallness of the wave-length of light the sides of the two dotted squares lie close to the edge of the aperture.

We have thus proved the rectilinear propagation of light on the basis of the wave theory, but at the same time have shown, as was mentioned on p. 3, that it is only approximate. The rays pass through the aperture unaffected only if they do not come too near the edge. Thus to the original statement of Huygens' principle Fresnel added the idea of the destructive interference of the different zones and explained why generally light waves do not bend round corners to any large extent. If an aperture in a screen were so small that the whole area lay in a single zone for all points on one side, the light rays though incident normally on the other side would diffuse in all directions on passing through. But for this to happen in the case of light the aperture would have to be very small indeed. In sound, however, it is the usual case. If p is 100 cms. and the wave-length that of Na light, the radius of the first zone is $\sqrt{p\lambda} = \cdot 076$ cms. In the corresponding problem in sound, if the wave-length under consideration is that of middle C on the piano, which is about 120 cms., the radius of the first zone is 109·4 cms. We can appreciate, therefore, what a difference the larger wave-length makes.

In this section only the case of a plane wave-front has been treated, but a spherical wave-front can be divided into zones and dealt with in the same way.

Mathematical Statement of Huygens' Principle. As mentioned on p. 155 the distance from P of points in the first zone varies between p and $p + \lambda/2$. By dividing the first zone into rings it can be shown that the phase of the resultant wave due to it at P is that of a wave that has travelled a distance $p + \lambda/4$ from O. But the actual wave travels only a distance p between O and P. Hence Huygens' principle, as extended by Fresnel, gives the phase of the wave wrong. It also does not explain why a wave is not propagated backwards.

This has been done by Kirchhoff, who has derived a more rigorous statement of the principle from the differential equation for wave-motion. This more rigorous statement also gives the phase right. A simplified form of its proof is given in Drude's "Optics". Here the result will merely be stated. It runs

$$4\pi s_0 = \iint\left(\frac{\partial}{\partial r}\left\{\frac{s(t - r/v)}{r}\right\}\cos(nr) - \frac{1}{r}\frac{\partial s(t - r/v)}{\partial n}\right)dS.$$

s_0 gives the light vibration at the point to be investigated. This point is taken as origin and round it a closed surface is drawn; dS is an element of area on this surface, n is the direction of the normal to dS drawn inwards, and r is the distance of dS from the origin. s represents the light vibration on dS as a function of t; before differentiation $t - r/v$ is substituted for t.

Diffraction. As has been shown, then, if a plane light-wave comes through an opening, it forms an image of that opening on a screen behind it, but the edges of the image are not sharp. Similarly, if a shadow is cast by an obstacle, the edges of the shadow are never perfectly sharp, no matter how parallel the incident light is. The preceding calculation does not inform us how the intensity varies at the edge, but we may anticipate the results of the following sections by stating, that the light always encroaches to some extent on the geometrical shadow, and that in the light near the edge of the beam there is a rhythmic variation in the brightness. If, for example, a very narrow slit is parallel to the straight edge of an obstacle and a screen is placed on the opposite side of the obstacle to receive its shadow, then the geometrical shadow is bounded by the plane through the slit touching the straight edge, and it is found that some light from the slit bends round behind the obstacle and meets the screen in the geometrical shadow, also immediately outside the geometrical shadow and parallel to its edge there are several dark bands.

Such phenomena are said to be due to diffraction. They were first studied by Grimaldi about the middle of the seventeenth century and afterwards by Hooke and Newton. The first attempt to explain them on the wave-theory was made by Young, who attributed them to the interference of the direct light that passed close to the edge with the light reflected at grazing incidence. If this explanation were true, the bands should be affected by the sharpness of the edge, its degree of polish and its material. Fresnel showed experimentally that they were not, and that hence Young's explanation is untenable. Diffraction phenomena are due to the interference of the direct light with itself. They are more difficult to produce and measure than interference bands.

In geometrical optics diffraction is ignored. Thus the laws of geometrical optics are not absolutely true. Even if the lens were perfect, a point image would never be formed of a point object but the image would always have a finite size.

Circular Aperture. If a plane wave is passing through a small circular aperture in a screen in a direction at right angles to the plane of the screen, the intensity can be determined for a point on the axis by dividing the aperture into zones. If there is an even number of zones in the aperture, the intensity is zero; if there is an odd number, all the zones cut out except the first one. In the ordinary propagation of a plane wave the amplitude at a point is half what would be caused by the first zone; hence in this case the amplitude and intensity have respectively twice and four times their values for an unlimited wave. If r is the radius of the aperture and b the distance of the point from the centre of the aperture, its distance from the circumference is $\sqrt{b^2 + r^2}$, which is equal to $b\sqrt{1 + r^2/b^2}$ or $b + r^2/(2b)$, since r is small in comparison with b. Hence for an even number of zones

$$\frac{r^2}{2b} = n\lambda \text{ and for an odd number } \frac{r^2}{2b} = (n + \tfrac{1}{2})\lambda,$$

where n is an integer. Consequently as the point moves in along the axis towards the aperture and $r^2/(2b)$ increases, the number of zones becomes odd and even in succession, and the intensity alternates between four times the normal and zero.

If the point is not on the axis, the edge of the aperture is no longer concentric with the zones and there is no elementary way of calculating the intensity. But in this case, if the aperture is small, by drawing it and the zones on a very large scale the intensity may be obtained graphically. For the sums of the areas of the odd and even zones included in the aperture can be measured either with a planimeter or by drawing them on squared paper and counting the squares, and the effect of the whole aperture may be regarded as proportional to their difference. By proceeding in this way we find that, if the aperture comprises only a few zones, the axis is surrounded by rings.

If the aperture is so small that it comprises only a fraction of the first zone, the point under investigation can be moved a considerable distance from the axis before the difference of its distances from the nearest and farthest points on the aperture amounts to half a wavelength. Consequently in this case the light diffuses far outside the geometrical shadow of the aperture.

It is obvious that in the case of the pinhole camera there is nothing to be gained in sharpness by making the hole smaller than the first half zone with reference to a point on the axis.

If instead of coming from infinity the incident wave diverges from a point on the axis distant a from the plane of the aperture, the wave-front is spherical and the condition for an even number of zones in the aperture is no longer $EP - CP = n\lambda$ but $EP - OP = n\lambda$.

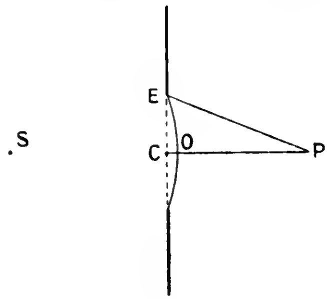


FIG. 141.

Since $CO = EC^2/2SO = r^2/2a$ approximately, this condition reduces to

$$\left(b + \frac{r^2}{2b}\right) - \left(b - \frac{r^2}{2a}\right) = \frac{r^2}{2}\left(\frac{1}{b} + \frac{1}{a}\right) = n\lambda.$$

Circular Obstacle. If a plane wave is incident on a circular disc in a direction at right angles to its plane and it is desired to find the intensity at points on its axis, we lay off zones in the plane of the disc taking the edge of the disc as the inner boundary of the first zone. Then, if the effects of the different zones are summed in the manner of p. 156, it is found that the effect of the whole wave is equal to half the effect of the first zone. If the disc is small, the latter has almost the same value as if there were no disc and it were laid off at the foot of the perpendicular. Hence the intensity on the axis is the same as if the disc were away. Round this bright spot on the axis there are alternately dark and bright rings.

The occurrence of the bright spot on the axis was deduced by Poisson as a result of Fresnel's reasoning and urged as an argument against the correctness of the latter. It was tested experimentally by Arago and he found the theory right. It had been discovered experimentally by Delisle about 1715, but his observations had been forgotten.

The experiment can be performed with a threepenny piece suspended vertically by threads and illuminated by sunlight through a pinhole at a distance of 15 or 20 feet. The bright spot should be viewed with an eyepiece at an equally great distance on the other side.

Zone Plates. On p. 155 the radii of the edges of the zones were given by

$$r^2_n = \left(p + \frac{n\lambda}{2}\right)^2 - p^2 = pn\lambda.$$

Draw on a plane screen a series of concentric circles with radii given by the above formula and suppose that these zones are alternately opaque and transparent. If a plane wave falls normally on the screen, the phases of the secondary waves emitted from each transparent zone agree at a point on the axis at a distance of p from the screen. If m denotes the effect of the first zone and N the total number of zones, then the amplitude at this point is roughly proportional to $\frac{1}{2}Nm$. The illumination at the point is thus roughly N times what it would be if the screen were removed. The screen thus acts as a convex lens of focal length p.

Such screens are called zone plates. They may be made by drawing concentric circles on a sheet of paper with their radii accurately proportional to the square roots of the first n natural numbers, blacking out the alternate zones and making a very reduced copy by photography. The glass negative will then act as a zone plate provided that the thickness of the glass is sufficiently uniform. It is, of course, extremely important that the lens used for the reproduction should have no appreciable distortion, i.e. it should not alter the relative lengths of the different radii. Fig. 142 is a copy of a drawing

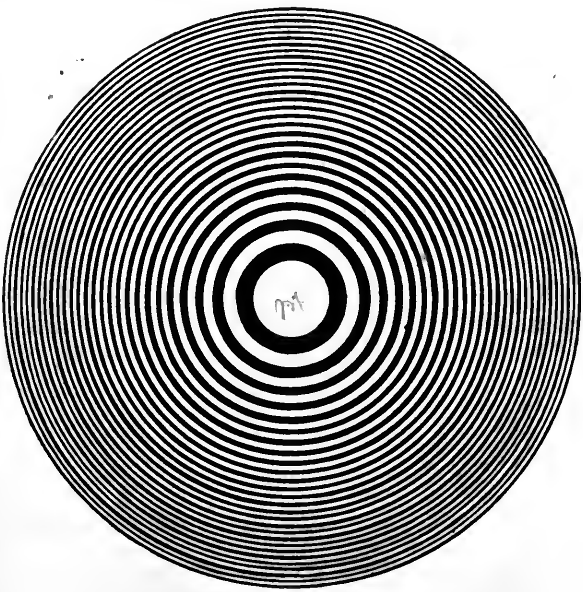


FIG. 142.

made for the purpose of reproducing zone plates from. They would have to be very much reduced; with the circles the same size as in the figure p would be about fifty metres.

There are formulæ for a zone plate analogous to those for a lens.

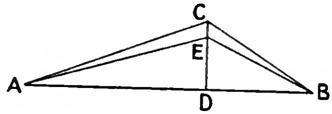


FIG. 143.

Let CD be a zone plate, AB be its axis, and let C and E be corresponding points on two successive clear zones. Since CD is small in comparison with AD we may write

$$AC = \sqrt{AD^2 + CD^2} = AD\sqrt{1 + CD^2/AD^2} = AD + CD^2/(2AD).$$

A corresponding expression may be obtained for CB ; hence

$$AC + CB = AD + DB + \frac{CD^2}{2}\left(\frac{1}{AD} + \frac{1}{DB}\right).$$

Similarly

$$AE + EB = AD + DB + \frac{ED^2}{2}\left(\frac{1}{AD} + \frac{1}{DB}\right).$$

The difference of the two paths from A to B is

$$\frac{1}{2}\left(\frac{1}{AD} + \frac{1}{DB}\right)(CD^2 - ED^2) = \frac{1}{2}\left(\frac{1}{AD} + \frac{1}{DB}\right)2p\lambda.$$

If $\frac{1}{AD} + \frac{1}{DB} = \frac{n}{p}$, the difference is $n\lambda$. Consequently the rays through any two successive transparent zones reinforce. Now n may have any integral value ; hence, if a luminous point is situated on the axis at a distance AD from the plate, images are formed of it on the axis at points B given by

$$\frac{1}{AD} + \frac{1}{DB} = \frac{n}{p}.$$

A zone plate thus differs from a lens in having several foci.

Let us suppose that instead of a point object at A there is a finite object of length AF (fig. 144). Let AF $= x$. Draw BG perpendicular

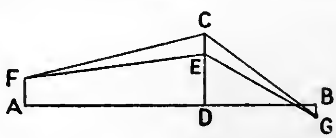


FIG. 144.

to AB, and let BG $= y$. Then

$$FC + CG = AD + DB + \frac{(CD - x)^2}{2AD} + \frac{(CD + y)^2}{2DB}$$

and

$$FE + EG = AD + DB + \frac{(ED - x)^2}{2AD} + \frac{(ED + y)^2}{2DB}.$$

The difference of the right hand sides is

$$\frac{1}{2AD}(CD^2 - ED^2 - 2xCE) + \frac{1}{2DB}(CD^2 - ED^2 + 2yCE)$$
$$= \frac{1}{2}(CD^2 - ED^2)\left(\frac{1}{AD} + \frac{1}{DB}\right) - CE\left(\frac{x}{AD} - \frac{y}{DB}\right).$$

If we assume that the first term equals $n\lambda$ and the second term equals zero, B is the image of A, and G is also the image of F. Thus the zone plate forms images of small extended objects and the magnification y/x follows the same law as in the case of a lens.

Cylindrical Wave-Front. Suppose AB is a cylindrical wave-front diverging from an infinitely long narrow slit at F and that it is required to investigate the effect of the wave-front at P. Let FP meet the cylinder in O. Let a be the radius of the cylinder and let OP $= b$.

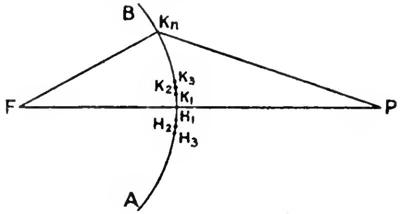


FIG. 145.

With centre P and radii $b + \frac{1}{2}\lambda$, $b + \lambda$, $b + \frac{3}{2}\lambda$, etc., draw arcs to cut the arc AB in K_1, H_1, K_2, H_2, K_3, H_3, etc. Let $PK_n = b + \frac{1}{2}n\lambda$. Then

$$PK_n^2 = FK_n^2 + FP^2 - 2FK_n\, FP \cos K_nFP,$$

or

$$\left(b + \frac{n\lambda}{2}\right)^2 = a^2 + (a + b)^2 - 2a(a + b) \cos K_nFP \quad . \quad (6)$$

Let $\angle K_nFP = \theta$. Then, if θ is small, $\cos K_nFP = 1 - \frac{1}{2}\theta^2$ and $n^2\lambda^2$ can be neglected in comparison with the other quantities. If we make these substitutions in (6) it becomes

$$b^2 + bn\lambda = 2a^2 + 2ab + b^2 - 2a(a + b)(1 - \tfrac{1}{2}\theta^2),$$

i.e.

$$bn\lambda = a(a + b)\theta^2,$$

or

$$\theta = \sqrt{\frac{nb\lambda}{a(a + b)}}.$$

The lengths of the arcs OK_1, K_1K_2, K_2K_3, etc., decrease in the ratio, 1, $\sqrt{2} - 1$, $\sqrt{3} - \sqrt{2}$, etc., or 1, ·414, ·318, very rapidly at first but more slowly afterwards.

If lines are drawn through K_1, K_2 . . . K_n parallel to the axis of the cylinder, they divide the surface into a series of strips. The different points on the surface of each strip are at different distances from P. For example, let fig. 146 represent a section of the cylinder by a plane through its axis and the point P. Then we can draw lines $PM_1 = b + \frac{1}{2}\lambda$, $PM_2 = b + \lambda$, $PM_3 = b + \frac{3}{2}\lambda$ to meet OM_3 in M_1, M_2, and M_3, i.e. we can divide the first strip into half-period elements. The other strips can be treated in the same way. It can be shown that the areas of these half-period elements decrease rapidly at first but more slowly afterwards. The effect of each strip reduces to that of the first few half-period elements, because the higher terms of the series representing its effect annul one another. Of course, the effect of the whole strip has the same sign as the effect of its first half-period element.

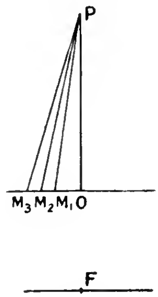


FIG. 146.

If we consider all the strips, now, we see that their effect at P can be represented by a series in which the odd and even terms have different signs and in which the size of the terms diminishes rapidly at first but slowly afterwards. The higher terms annul one another, and the effect of the whole series is equal to the effect of the first few terms. The effects of the first few terms are not equal as on p. 155 where the zones had equal areas.

Elementary Theory of Diffraction at a Straight Edge. Let F be a long narrow slit from which a cylindrical wave diverges. Let TV be a thin metal plate with a straight edge at T parallel to the slit and let PM be a screen. Then, if the straight line FT be produced to meet the screen in M, M is the edge of the geometrical shadow. According to the laws of geometrical optics the portion ML of the screen should be wholly dark and the portion MP should be illuminated and of uniform intensity. The illumination should start discontinuously at M. As has been mentioned on p. 159, the laws of geometrical optics are not absolutely true.

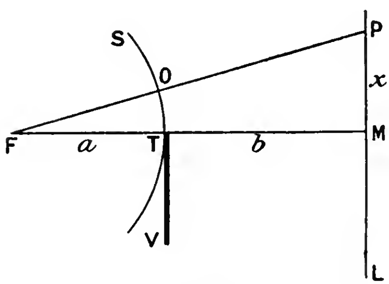


FIG. 147.

Let FT = a, let TM = b, and let MP = x. Let P be any point on the screen in the plane of the diagram. Draw a cylinder with the slit as axis to touch the straight edge at T; join FP and let it meet this cylinder at O.

With P as centre and radii PO + $\frac{1}{2}\lambda$, PO + λ, PO + $\frac{3}{2}\lambda$, etc., mark off points on the wave-front, draw lines through these points parallel to the axis of the cylinder and so divide the surface of the cylinder into a series of strips. The series on the side OS is complete; hence that part of the wave-front always gives at P half the amplitude that would be given by a complete wave.

The series on the side OT is incomplete, the higher members being obscured by the plate TV. If there is only one strip clear, the amplitude due to this side is large; if there are two strips clear, they almost annul one another and the amplitude due to this side is small.

If there are three strips clear we have a maximum again. If there is an odd number of strips in OT, the amplitude at P is a maximum, and if there is an even number of strips in OT, it is a minimum.

If x is small compared with b,

$$\text{PT} = \sqrt{b^2 + x^2} = b\left(1 + \frac{x^2}{b^2}\right)^{\frac{1}{2}} = b\left(1 + \frac{x^2}{2b^2}\right) = b + \frac{x^2}{2b}.$$

Similarly

$$\text{PF} = \sqrt{(a + b)^2 + x^2} = a + b + \frac{x^2}{2(a + b)}$$

and hence

$$\text{PO} = b + \frac{x^2}{2(a + b)}.$$

For a minimum at P we have PT − PO = $n\lambda$ where n is any integer. This gives

$$\left(b + \frac{x^2}{2b}\right) - \left(b + \frac{x^2}{2(a + b)}\right) = n\lambda,$$

$$\frac{x^2}{2}\left(\frac{1}{b} - \frac{1}{a + b}\right) = n\lambda,$$

$$x^2 \frac{a}{2b(a + b)} = n\lambda,$$

and finally

$$x = \sqrt{\frac{b(a + b)\, 2n\lambda}{a}}:$$

Similarly for a maximum at P we have

$$x = \sqrt{\frac{b(a + b)\,(2n + 1)\lambda}{a}}.$$

Thus, as the point P moves upwards from M, it passes through maxima and minima of illumination. Consequently there are bright and dark bands on the screen parallel to the edge of the geometrical shadow. As the effects of the first few strips are by no means equal the above formulæ are not very accurate and cannot be relied on at all unless n is small. If P is on the edge of the geometrical shadow, the amplitude is half what would be produced by a complete wave, i.e. half what would be produced on the screen at a point far out from M, and the intensity is one-quarter of its value there.

If P is in the geometrical shadow, all the strips are obscured on one side of O and the first strip on the other side starts at some distance from O. It is found then on considering the numerical value, that the intensity diminishes steadily from M into the geometrical shadow and becomes zero at a short distance from M.

Diffraction at a Straight Edge. Fresnel's Theory. The figure is the same as fig. 147 except that a straight line has been drawn through P to meet OS in Q. Let OQ be s, measured along the arc, let OP = c, and let PQ = $c + \delta$. Then, if s is small, we can regard OQ as perpendicular to OP, and $\text{PQ}^2 = \text{OP}^2 + \text{OQ}^2$ or $(c + \delta)^2 = s^2 + c^2$.

Since δ is small the δ^2 term can be neglected in this equation and we obtain

$$\delta = \frac{s^2}{2c}.$$

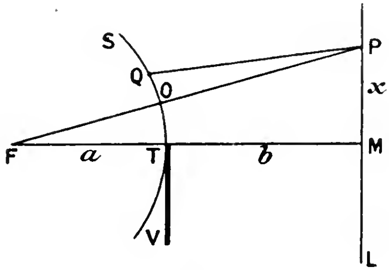


FIG. 148.

Let $\sin 2\pi t/\tau$ represent the vibration on the cylindrical surface. Every point on the surface may be regarded as a secondary source sending out waves to P. If we consider a strip of width ds at Q, the vibration due to it at P is proportional to

$$\sin 2\pi\left(\frac{t}{\tau} - \frac{c+\delta}{\lambda}\right)ds \quad . \quad . \quad . \quad . \quad . \quad (7)$$

If we write OT $= -s_0$ and OS $= s_1$, the resultant vibration at P due to the whole wave front may be written

$$\int_{-s_0}^{s_1} \sin 2\pi\left(\frac{t}{\tau} - \frac{c+\delta}{\lambda}\right)ds \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The amplitude at P should depend also on the distance of the secondary source, and to be perfectly accurate the expression (7) should be divided by 1/PQ. Fresnel makes the assumption that it is only the strips in the neighbourhood of O that matter, that for these PQ can be considered constant, and that the other strips can be neglected. This assumption is justified by his results.

Since δ is the only quantity in (8) depending on s, the integral may be written

$$\sin 2\pi\left(\frac{t}{\tau} - \frac{c}{\lambda}\right)\int_{-s_0}^{s_1} \cos\frac{2\pi\delta}{\lambda}\,ds - \cos 2\pi\left(\frac{t}{\tau} - \frac{c}{\lambda}\right)\int_{-s_0}^{s_1} \sin\frac{2\pi\delta}{\lambda}\,ds,$$

and reduces to

$$\mathrm{R}\sin\left\{2\pi\left(\frac{t}{\tau} - \frac{c}{\lambda}\right) - \theta\right\},$$

if we write $\mathrm{R}\cos\theta = \int_{-s_0}^{s_1} \cos\frac{2\pi\delta}{\lambda}\,ds$, and $\mathrm{R}\sin\theta = \int_{-s_0}^{s_1} \sin\frac{2\pi\delta}{\lambda}\,ds.$

The intensity of the resultant vibration at P is therefore proportional to R^2, i.e. to

$$\left(\int_{-s_0}^{s_1} \cos \frac{2\pi\delta}{\lambda} ds\right)^2 + \left(\int_{-s_0}^{s_1} \sin \frac{2\pi\delta}{\lambda} ds\right)^2.$$

We have already found that $\delta = s^2/(2c)$. Introduce a new variable v such that

$$s = \sqrt{\frac{c\lambda}{2}}\, v.$$

Then $\dfrac{2\pi\delta}{\lambda} = \dfrac{\pi v^2}{2}$ and $ds = \sqrt{\dfrac{c\lambda}{2}}\, dv$. When $s = s_0$, let $v = v_0$. When $s = s_1$, $v = s_1 \Big/ \sqrt{\dfrac{c\lambda}{2}}$; s_1 is finite, λ is very small, consequently the value of v corresponding to the upper limit of integration is very great, and it is found that without influencing the results it may be put $= +\infty$.

We find finally therefore that the intensity at P is proportional to

$$\left(\int_{-v_0}^{+\infty} \cos \frac{\pi v^2}{2} dv\right)^2 + \left(\int_{-v_0}^{+\infty} \sin \frac{\pi v^2}{2} dv\right)^2 \quad . \quad . \quad (9)$$

The integrals inside the brackets are known as Fresnel's integrals. They have been evaluated in the form of series, and tables have been drawn up giving their numerical values for different values of v. From these tables the values of (9) can be obtained for different positions of P. There is, however, an elegant geometrical way, due to Cornu, of studying the variation of (9).

Cornu's Spiral. Cornu's spiral is defined by

$$x = \int_0^v \cos \frac{\pi v^2}{2} dv, \quad y = \int_0^v \sin \frac{\pi v^2}{2} dv \quad . \quad . \quad (10)$$

When $v = 0$, $x = y = 0$; consequently it passes through the origin, and, since $\sin \dfrac{\pi v^2}{2} = 0$ for $v = 0$, it touches the x axis there. From (10)

$$dx = \cos \frac{\pi v^2}{2} dv, \quad dy = \sin \frac{\pi v^2}{2} dv,$$

hence $\dfrac{dy}{dx} = \tan \dfrac{\pi v^2}{2}$, and ψ, the angle which the tangent makes with the x axis, is given by

$$\psi = \frac{\pi v^2}{2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

We have

$$ds = \left(1 + \left(\frac{dy}{dx}\right)^2\right)^{\frac{1}{2}} dx = \left(1 + \tan^2 \frac{\pi v^2}{2}\right)^{\frac{1}{2}} \cos \frac{\pi v^2}{2} dv = dv;$$

hence $s = v$, and combining this with (11) we find for the intrinsic equation of the curve

$$\psi = \frac{\pi s^2}{2}.$$

We have $d\psi = \pi s ds$; hence the curvature is given by

$$\frac{1}{\rho} = \frac{d\psi}{ds} = \pi s.$$

At the origin it is zero and changes sign; consequently there is a point of inflection there. The curvature also increases continuously as we move along the curve from the origin both ways.

It can be shown that

$$\int_0^\infty \cos\frac{\pi v^2}{2}\,dv = \int_0^\infty \sin\frac{\pi v^2}{2}\,dv = \tfrac{1}{2},$$

hence the curve has asymptotic points at $(\frac{1}{2}, \frac{1}{2})$ and $(-\frac{1}{2}, -\frac{1}{2})$.

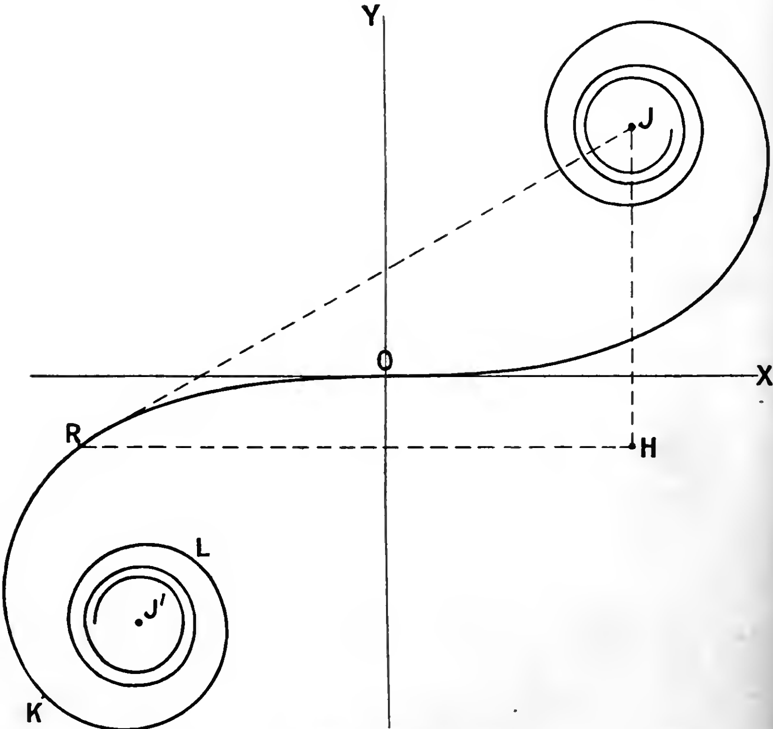


Fig. 149.—Cornu's Spiral.

In fig. 149 the curve has been drawn accurately from the values of the integrals given in the tables.

Application of Cornu's Spiral. If in fig. 149 we mark off OR $= -v_0$, and if through R and J, the asymptotic point, we draw straight lines parallel respectively to OX and OY, then

$$\text{RH} = \int_{-v_0}^{+\infty} \cos \frac{\pi v^2}{2}\, dv, \text{ and } \text{HJ} = \int_{-v_0}^{+\infty} \sin \frac{\pi v^2}{2}\, dv.$$

The expression proportional to the intensity on p. 167, namely (9), is therefore equal to

$$\text{RH}^2 + \text{HJ}^2 = \text{RJ}^2.$$

Let us suppose that we start with the point P in fig. 148 well into the shadow and gradually move it up the screen. When it is well into the shadow, $-v_0$ is $+\infty$, consequently the point R is at J and the intensity is zero as represented by the point A in fig. 150. As P moves to M in fig. 148, R moves from J to O on the spiral. During this change RJ^2 steadily increases, and when P reaches M the intensity can be represented by some such point as B in fig. 150. When P passes above M in fig. 148 the point crosses over on to the other part of the spiral and RJ^2 increases until R reaches the point K; this corresponds to the maximum C on the intensity curve. As the point R moves from K to L on the spiral, RJ^2 diminishes; the point L on the spiral corresponds to the minimum D on the intensity curve. And so on: as R runs round the convolutions of the spiral, it gives rise to a series of gradually decreasing undulations on the intensity curve until finally, when R reaches the asymptotic point J′, the intensity curve has become a straight line. As $\text{JJ}'^2 = 4\text{JO}^2$, the ordinate of this region of the intensity curve is four times as high as the ordinate at B, which corresponds to the edge of the geometrical shadow.

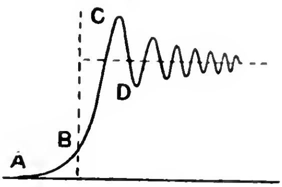


Fig. 150.

Other diffraction problems can be treated in a similar way by the application of the spiral. For example, in the case of diffraction at a narrow slit the intensity varies as the square of the chord joining two points on the spiral, which are at a constant distance apart measured along the arc. This constant distance is proportional to the width of the slit.

The Composition of Simple Harmonic Motions. Let us suppose that at the same time a particle is executing two different simple harmonic motions in the same straight line, and that these simple harmonic motions have the same period but different amplitudes and different phases. Then its resultant displacement may be represented by

$$\begin{aligned} y &= a_1 \cos(\omega t - \alpha_1) + a_2 \cos(\omega t - \alpha_2) \\ &= a_1 \cos \omega t \cos \alpha_1 + a_1 \sin \omega t \sin \alpha_1 + a_2 \cos \omega t \cos \alpha_2 + a_2 \sin \omega t \sin \alpha_2 \\ &= (a_1 \cos \alpha_1 + a_2 \cos \alpha_2) \cos \omega t + (a_1 \sin \alpha_1 + a_2 \sin \alpha_2) \sin \omega t. \end{aligned}$$

Write $a_1 \cos \alpha_1 + a_2 \cos \alpha_2 = \text{R} \cos \delta$, and $a_1 \sin \alpha_1 + a_2 \sin \alpha_2 = \text{R} \sin \delta$. Then

$$\begin{aligned} y &= \text{R} \cos \delta \cos \omega t + \text{R} \sin \delta \sin \omega t \\ &= \text{R} \cos(\omega t - \delta). \end{aligned}$$

The two S.H.Ms. thus combine into one S.H.M. of the same period, the amplitude and phase of the resultant S.H.M. being given by

$$R^2 = (a_1 \cos \alpha_1 + a_2 \cos \alpha_2)^2 + (a_1 \sin \alpha_1 + a_2 \sin \alpha_2)^2$$
$$= a_1^2 + a_2^2 + 2a_1a_2 \cos(\alpha_1 - \alpha_2)$$

and
$$\tan \delta = \frac{a_1 \sin \alpha_1 + a_2 \sin \alpha_2}{a_1 \cos \alpha_1 + a_2 \cos \alpha_2}.$$

Let $\angle AOP = \alpha_1$ and $OP = a_1$, let $\angle AOQ = \alpha_2$ and $OQ = a_2$, and complete the parallelogram POQS. Then $\cos(\alpha_1 - \alpha_2)$ is obviously equal to $\cos QOP = -\cos OPS$, and

$R^2 = a_1^2 + a_2^2 + 2a_1a_2 \cos(\alpha_1 - \alpha_2) = OP^2 + PS^2 - 2OP \cdot PS \cos OPS = OS^2$.
Draw PN and SM perpendicular to OA. Then

$ON = a_1 \cos \alpha_1$, $NM = a_2 \cos \alpha_2$, and $OM = a_1 \cos \alpha_1 + a_2 \cos \alpha_2$.
Similarly $MS = a_1 \sin \alpha_1 + a_2 \sin \alpha_2$. Thus

$$\tan SOA = \frac{MS}{OM} = \frac{a_1 \sin \alpha_1 + a_2 \sin \alpha_2}{a_1 \cos \alpha_1 + a_2 \cos \alpha_2} = \tan \delta$$

and $\angle SOA = \delta$. That this is the only solution may be seen by considering the limiting case when $a_2 = 0$.

Hence, if OP and OQ represent the amplitudes and phases of the component S.H.Ms., OS represents the amplitude and phase of the resultant. We thus arrive at the result that S.H.Ms. of the same period can be compounded by the parallelogram construction, or, if there are more than two of them, by the polygon construction, exactly in the same way as forces or velocities.

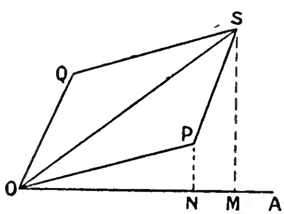


FIG. 151.

This throws a new light on the summation of the series on p. 156. If the first zone is divided into nine smaller zones of equal area, the amplitudes of the vibrations caused by each of these are equal and the phases may be represented by angles of 10°, 30°, 50°, 70°, 90°, 110°, 130°, 150°, and 170°; consequently the effect of the whole zone is represented by the unclosed polygon in fig. 152. In the limit when the

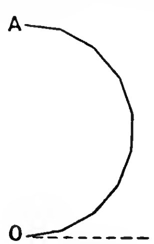


FIG. 152.

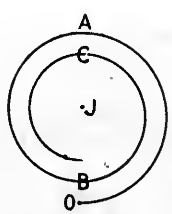


FIG. 153.

first zone is divided into an infinite number of smaller zones, the unclosed polygon becomes the semicircle OA in fig. 153. In the same way the effect of the second zone may be represented by a second semicircle AB of slightly smaller radius, because the second zone is at a slightly greater distance; the third zone may be represented by the still smaller semicircle BC and the whole wave-front represented by a spiral converging to the asymptotic point J. By the construction just proved, the effect of the first zone may be represented by the diameter OA and the effect of the whole spiral by the straight line joining OJ. Hence the effect of the whole spiral is equal to half the effect of the first zone.

n S.H.Ms. with Equal Amplitude, Same Period and Phases increasing in an Arithmetical Progression. Let a be the common amplitude and d the common phase difference. Let R and δ be the amplitude and phase of the resultant; then, by the polygon construction, if we resolve parallel and perpendicular to the first side,

$$\text{R}\cos\delta = a(1 + \cos d + \cos 2d \ldots + \cos\{n-1\}d),$$
$$\text{R}\sin\delta = a(\sin d + \sin 2d \ldots + \sin\{n-1\}d).$$

If we multiply both of these series by $2\sin\frac{d}{2}$, the first gives

$$2\text{R}\cos\delta\sin\frac{d}{2} = a\Big(2\sin\frac{d}{2} + 2\cos d\sin\frac{d}{2} + 2\cos 2d\sin\frac{d}{2} \ldots + 2\cos\{n-1\}d\sin\frac{d}{2}\Big)$$
$$= a\Big(2\sin\frac{d}{2} + \sin\frac{3d}{2} - \sin\frac{d}{2} + \sin\frac{5d}{2} - \sin\frac{3d}{2} \ldots + \sin\{n-\tfrac{1}{2}\}d - \sin\{n-\tfrac{3}{2}\}d\Big)$$
$$= a\Big(\sin\frac{d}{2} + \sin\{n-\tfrac{1}{2}\}d\Big)$$

i.e. $$\text{R}\cos\delta\sin\frac{d}{2} = a\sin\frac{nd}{2}\cos\frac{(n-1)d}{2} \quad . \quad . \quad . \quad . \quad (12)$$

Similarly the second gives

$$\text{R}\sin\delta\sin\frac{d}{2} = a\sin\frac{nd}{2}\sin\frac{(n-1)d}{2} \quad . \quad . \quad (13)$$

On squaring (12) and (13) and adding we obtain

$$\text{R}^2\sin^2\frac{d}{2} = a^2\sin^2\frac{nd}{2} \text{ or } \text{R} = \frac{a\sin\frac{nd}{2}}{\sin\frac{d}{2}},$$

and on dividing (13) by (12) we obtain

$$\tan\delta = \tan\frac{(n-1)d}{2} \text{ or } \delta = \frac{(n-1)d}{2}.$$

If we take the other solution, $\delta = \frac{(n-1)d}{2} + \pi$, consideration of the limiting case when $n = 1$ shows that we must then choose the other sign in taking the root of R^2 and consequently get the same result over again.

Suppose, now, that n becomes infinitely great but that na is kept finite and $= A$ and nd finite and $= 2\alpha$. Then

$$R = \frac{a \sin \alpha}{\sin \alpha/n}$$
$$= \frac{a \sin \alpha}{\alpha/n}, \text{ since } \alpha/n \text{ is small,}$$
$$= \frac{A \sin \alpha}{\alpha}$$
$$\text{and } \delta = \alpha,$$

since in the limit $n - 1$ is equal to n. This latter result may be obtained very easily graphically. The unclosed polygon becomes in the limiting case an arc of a circle of length A, which subtends an angle 2α at the centre. The chord is, of course, parallel to the tangent at the mid point of the arc, i.e. $\delta = \alpha$, and the length of the chord is twice the radius multiplied by $\sin \alpha$, i.e.

$$2\left(\frac{A}{2\alpha}\right) \sin \alpha \quad \text{or} \quad \frac{A \sin \alpha}{\alpha}.$$

Passage of a Plane Wave through a Slit. Let AB be the slit, let e be its breadth, and let us suppose first that the wave is incident in a direction at right angles to its plane. We shall suppose the slit to be infinitely long; then as conditions do not vary in a direction perpendicular to the plane of the figure we can restrict ourselves to the plane of the figure. Each point in AB may be regarded as a secondary source sending out waves in all directions. In the diagram are drawn the rays emitted from A and B in a direction making an angle θ with the normal. Let us find the resultant of all the rays emitted in this direction. Draw AK perpendicular to BK.

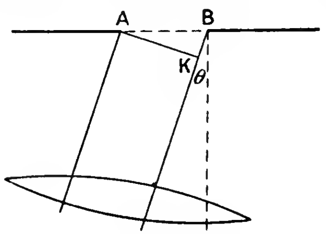


FIG. 154.

The rays may be supposed either collected by a lens, in which case they meet in a point in the focal plane and the lengths of the paths of the different rays from AK to this point are the same, or they may be supposed falling on a screen at infinity, in which case all lines from AK to a point in the direction θ are equal, because they are inclined at an infinitely small angle to BK.

The path differences of the different rays therefore vary uniformly from 0 to BK, i.e. from 0 to $e \sin \theta$, and the phase differences from 0 to $(2\pi e \sin \theta)/\lambda$. Applying the result of the preceding section we find that the amplitude is

$$\frac{A \sin \alpha}{\alpha} \text{ where } \alpha = \frac{\pi e \sin \theta}{\lambda}.$$

To find the turning values of this expression differentiate with respect to a and equate to zero. Then

$$\frac{d}{da}\frac{\sin a}{a} = \frac{\cos a}{a} - \frac{\sin a}{a^2} = 0$$

or $a - \tan a = 0$.

This equation can be solved graphically; if we draw the graphs of $\tan a$ and of a straight line making an angle of 45° with the axes, the abscissæ of the points of intersection are the roots of the equation. By inspection of fig. 155 we find that the first root is 0, then we have a series of roots less than and gradually getting closer to $\frac{3}{2}\pi$, $\frac{5}{2}\pi$, $\frac{7}{2}\pi$, etc., respectively. The exact values of the first seven are $1{\cdot}430\pi$, $2{\cdot}459\pi$, $3{\cdot}471\pi$, $4{\cdot}477\pi$, $5{\cdot}482\pi$, $6{\cdot}484\pi$, and $7{\cdot}486\pi$. These roots

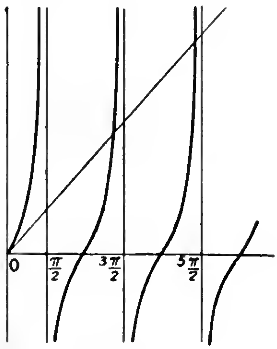


FIG. 155.

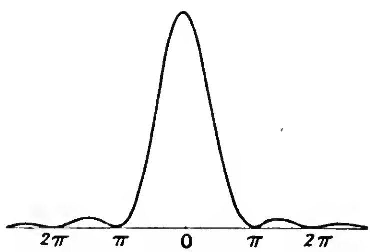


FIG. 156.

though giving both the maxima and minima of the amplitude give the maxima alone of the intensity, for the latter is equal to

$$\frac{A^2 \sin^2 a}{a^2}$$

and cannot be negative. The minima of intensity are obviously given by $a = n\pi$, where n has any integral value except 0; the latter, of course, gives a maximum. For $a = 0$ the intensity is equal to A^2. Fig. 156 represents the intensity as a function of a. The maxima are not equidistant from adjacent minima but lie nearer the centre. The heights of the second, third, and fourth maxima are respectively $\frac{1}{21}$, $\frac{1}{61}$, and $\frac{1}{120}$ if A is put = 1. The area of the curve $= 2\int_0^\infty \frac{\sin^2 a}{a^2}\, da = \pi$. The positions of the first minima are given in terms of θ by $e \sin\theta = \pm\lambda$; hence except in extreme cases $\sin\theta$ is small and θ can be taken proportional to a. If e is made comparable with λ, the system of bands widens out and light spreads out in all directions from the slit.

The diffraction pattern discussed in this section can easily be observed with an ordinary spectrometer. The instrument is first focussed for parallel light in the usual way, illuminated with sodium light, and then a narrow slit mounted on the prism table. When the collimator slit has been set as parallel to this slit as possible, three or four maxima can be seen on each side of the central maximum.

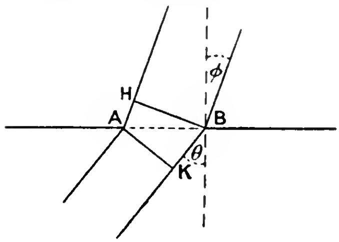


FIG. 157.

If instead of being incident normally the incident rays make an angle ϕ with the normal to the plane of the slit (fig. 157), the path difference of the extreme rays is e (sin θ − sin ϕ) and the preceding discussion holds if we put

$$\alpha = \frac{\pi e(\sin\theta - \sin\phi)}{\lambda}.$$

As the central maximum is given by $\alpha = 0$, i.e. by $\sin\theta = \sin\phi$, if in the arrangement described above the prism table with the slit is rotated, the central maximum remains undeviated. At the same time the bands widen out, for the positions of the first minima are given by $e(\sin\theta - \sin\phi) = \pm\lambda$; hence if we write δ for the deviation and regard ϕ as fixed, $\theta = \phi + \delta$ and this equation gives

$$e(\sin(\phi + \delta) - \sin\phi) = \pm\lambda \quad \text{or} \quad e\cos\phi\sin\delta = \pm\lambda,$$

since δ is small; consequently $\sin\delta$ is increased by the rotation in the ratio of $\cos\phi$ to 1.

The Diffraction Grating. The diffraction grating is an instrument used for producing spectra. It consists usually of a number of fine, equidistant and parallel lines ruled with a diamond on a mirror of speculum metal or on a glass plate, or of a celluloid cast made from such a grating. Some of the earliest gratings consisted of fine wires stretched parallel and equidistant from one another between two screws of equal pitch. Gratings ruled with a diamond on a dividing machine are very expensive, hence celluloid casts or replicas are used instead, except for the most important work. Such replicas are made by T. Thorp, Manchester, and are consequently known as Thorp gratings. They are, of course, transparent.

Previous to 1883 all gratings had been plane; they had been ruled on plane surfaces or the wires had been stretched in one plane. They were used with telescope and collimator just as the prism is used in the spectroscope. But in that year Rowland described a method of ruling them on concave metal mirrors, which rendered the use of collimator and telescope lenses superfluous. In this section only plane gratings will be treated.

We can divide gratings into two classes, transmission gratings and reflection gratings. The first include wire gratings, those ruled on glass, and celluloid replicas; the second include all gratings ruled on metal.

In the elementary theory of the grating it is assumed to consist of a glass plate on which there is a great number of equidistant, parallel, opaque strips or lines of equal width. These lines are separated by clear spaces, A_1B_1, etc. Let A_1A_2, the width of a combined line and clear space, be e.

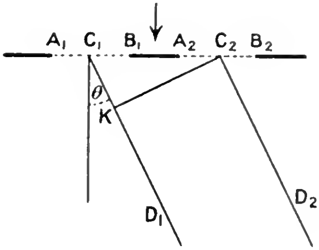


FIG. 158.

If a plane wave of wave-length λ is incident perpendicularly on the grating, all the clear spaces act as secondary sources and emit rays in all directions. Let us consider the rays diffracted at an $\angle\theta$ with the normal. Then all the rays from each space have a resultant, which has the same phase as the ray coming from the mid point of that space. We can consider them replaced by this resultant, and the problem reduces to that of compounding a number of parallel rays, C_1D_1, C_2D_2, etc., from a number of equidistant points, C_1, C_2, etc.

Draw C_2K perpendicular to C_1D_1. Then the common difference of these rays is $C_1K = C_1C_2 \sin C_1C_2K = e \sin\theta$, and the rays reinforce in the direction θ if

$$e \sin\theta = n\lambda,$$

where n has any integral value, positive or negative. If the telescope is pointed in a direction given by this equation, a bright line is seen on the cross-wires.

Most gratings have about 14,000 lines to the inch, in which case e is $\frac{1}{14000}$ inch or about $1{\cdot}81\ 10^{-4}$ cms. If the light used is yellow and of wave-length $5{\cdot}8\ 10^{-5}$ cms., the equation for the maxima becomes

$$\sin\theta = {\cdot}3197\, n$$

and the values of θ given by $n = 0, \pm 1, \pm 2, \pm 3$ are $0°, \pm 18°\ 39'$, $\pm 39°\ 45'$, $\pm 73°\ 33'$ respectively. Larger values of n give impossible values for $\sin\theta$. The image given by $\theta = 0$ is called the direct image, the two given by $n = \pm 1$ are said to belong to the first order spectrum, those given by $n = \pm 2$ are said to belong to the second order spectrum, and those given by $n = \pm 3$ to the third order spectrum.

If white light is incident on the grating, each colour into which it can be resolved forms its own images. The direct images all superimpose, hence the direct image formed is white, but the other images do not superimpose. The first order images form two first order spectra, the second order images two second order spectra, and the third order images two third order spectra, so that altogether this grating forms six spectra. These different spectra overlap. For example, if in the equation $e \sin\theta = n\lambda$ we write $e \sin\theta = 1{\cdot}5\ 10^{-4}$, a possible value, we find that $n = 2$ gives $7{\cdot}5\ 10^{-5}$ cms. and $n = 3$ gives $5\ 10^{-5}$ cms., so that the red of the second order spectrum is superimposed on the green of the third order spectrum.

If the light is incident at an $\angle\phi$ (fig. 159) instead of normally, the common path difference is

$$C_1K - C_2H = e(\sin\theta - \sin\phi),$$

and the maxima are given by

$$e(\sin\theta - \sin\phi) = n\lambda.$$

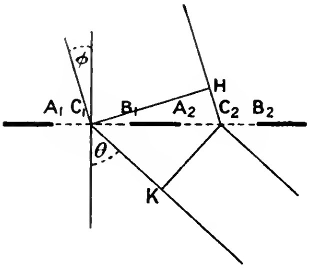


FIG. 159.

The deviation is given by $\theta - \phi$. For it to be a minimum we must have $d\theta - d\phi = 0$. Now, if the wave-length be regarded as fixed, we have

$$e(\cos\theta\, d\theta - \cos\phi\, d\phi) = 0,$$

which gives on substituting $d\theta = d\phi$

$$\cos\theta - \cos\phi = 0,$$

or $\theta = \pm\phi$. As $\theta = \phi$ obviously gives the direct image, $\theta = -\phi$ is the condition for minimum deviation. It requires the incident and emergent rays to make equal angles with the grating.

More Accurate Theory of Grating. The case treated in the previous section is an ideal one. No grating consists of a plane surface with perfectly transparent and perfectly opaque strips; the rulings in a transparent grating are always translucent and of an irregular form. This, however, makes no essential difference in the reasoning; it is not necessary for the rays to be stopped by the rulings, it is only necessary for the surface to have a periodic structure.

Let fig. 160 represent a section of a transmission grating which has altogether N rulings, and suppose that the light is incident perpendicularly on the plane side. Let e be the distance A_1A_2 measured along the chord. Consider the rays that are diffracted at an $\angle\theta$ from the ruling A_1A_2. If they are collected by a lens, the length of the path from the plane face of the grating to the image varies from ray to ray, but it is always possible to find some point C_1 between A_1 and A_2, such that the ray through it has the same phase at the image as the resultant. The rays from A_1A_2 can be regarded as equivalent to a single ray coming from C_1, the rays from A_2A_3 as equivalent to a single ray coming from the corresponding point C_2, and the rays from the whole grating as equivalent to N rays coming from N equidistant points arranged along a line. The common phase difference is $(2\pi e \sin\theta)/\lambda$ in radian measure, hence applying the result of p. 171 we find that the resultant intensity in the direction θ is proportional to

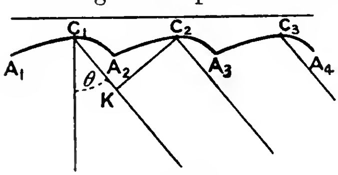


FIG. 160.

$$\left(\frac{\sin\dfrac{N\pi e\sin\theta}{\lambda}}{\sin\dfrac{\pi e\sin\theta}{\lambda}}\right)^2.$$

In order to study the variation of this expression with $\frac{\pi e \sin\theta}{\lambda}$ it is best to plot it as a function of the latter for a small value of N, say N = 6. When $\frac{\pi e \sin\theta}{\lambda} = n\pi$ the expression becomes indeterminate, but if we apply the usual rule and differentiate both numerator and denominator we find for the amplitude

$$\frac{\frac{d}{d\theta}\sin\frac{N\pi e \sin\theta}{\lambda}}{\frac{d}{d\theta}\sin\frac{\pi e \sin\theta}{\lambda}} = \frac{N\cos\frac{N\pi e\sin\theta}{\lambda}}{\cos\frac{\pi e \sin\theta}{\lambda}} = \frac{N\cos Nn\pi}{\cos n\pi} = \pm N,$$

hence at these points the intensity is proportional to N^2, or in the present case, to 36. Fig. 161 gives the graph; it has principal maxima at points given by $e \sin\theta = n\lambda$, and between every two principal maxima four much smaller subsidiary maxima. In the general case N has a large value, perhaps 14,000; between every two principal maxima there are $N-2$ subsidiary maxima, but they occur so close together and owing to the large value of N^2 they are so faint in comparison with the principal maxima that they cannot be observed. Only the principal maxima need to be considered; they are the only maxima given by the elementary theory on p. 175, and they are the images of the different orders met with in the ordinary use of the grating.

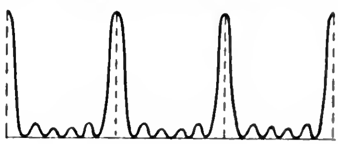


FIG. 161.

To study the variation of the amplitude in the neighbourhood of a maximum write $\theta + d\theta$ for θ, where θ is given by $e\sin\theta = n\lambda$; θ gives the position of the maximum and $d\theta$ the distance of the point in question from the maximum. The amplitude then becomes

$$\frac{\sin\frac{N\pi e}{\lambda}\sin(\theta + d\theta)}{\sin\frac{\pi e}{\lambda}\sin(\theta+d\theta)} \text{ or } \frac{\sin\frac{N\pi e}{\lambda}(\sin\theta + d\theta\cos\theta)}{\sin\frac{\pi e}{\lambda}(\sin\theta + d\theta\cos\theta)},$$

since $d\theta$ is small. This is equal to

$$\frac{\sin\left(Nn\pi + \frac{N\pi e}{\lambda}d\theta\cos\theta\right)}{\sin\left(n\pi + \frac{\pi e}{\lambda}d\theta\cos\theta\right)} = \pm\frac{\sin\left(\frac{N\pi e}{\lambda}d\theta\cos\theta\right)}{\sin\left(\frac{\pi e}{\lambda}d\theta\cos\theta\right)}.$$

Assume that the denominator is so small that the angle can be written for the sine, and substitute $n/\sin\theta$ for e/λ. Then the expression becomes

$$\pm\frac{\sin(Nn\pi\cot\theta\, d\theta)}{n\pi\cot\theta\, d\theta}.$$

But this is of the form

$$\frac{\sin \alpha}{\alpha}$$

and consequently has its first minimum at $\alpha = \pi$, i.e. at $\mathrm{N}n \cot \theta \; d\theta = 1$ or

$$d\theta = \frac{1}{\mathrm{N}n \cot \theta}.$$

This expression gives the angular distance from a principal maximum to the first point of zero intensity on either side of it. The value obtained for $d\theta$ obviously justifies the assumptions made above, namely, that $d\theta$ is small and that the angle can be written for the sine in the denominator of the expression for the amplitude.

Resolving Power of a Grating. If in a spectrum there are two lines very close together and their images partly superimpose, it is assumed, and is on the whole borne out by experience, that they can just be recognised as separate when the maximum of the one is on the first minimum of the other. If they are closer together than this they appear to be one single line.

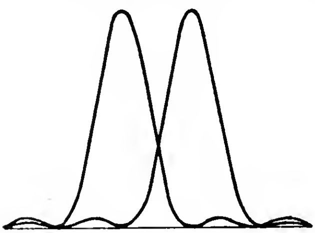


FIG. 162.

If λ is the wave-length of the one line and $\lambda + d\lambda$ the wave-length of the other, we have

$$n\lambda = e \sin \theta, \quad n d\lambda = e \cos \theta d\theta,$$

and consequently the angular separation between the two lines is given by $d\theta = \frac{n}{e \cos \theta} d\lambda$. But if this is the distance between the maximum of one line and its first minimum

$$d\theta = \frac{1}{\mathrm{N}n \cot \theta}.$$

Hence, on equating,

$$\frac{n d\lambda}{e \cos \theta} = \frac{1}{\mathrm{N}n \cot \theta} \text{ and } d\lambda = \frac{e \sin \theta}{\mathrm{N}n^2} = \frac{\lambda}{\mathrm{N}n},$$

which gives $\frac{\lambda}{d\lambda} = \mathrm{N}n$. This last expression is called the resolving power of the grating; it is the wave-length at any point in the spectrum divided by the least difference of wave-length that can be detected there, and it is equal to the product of the order of the spectrum and the total number of rulings in the grating.

Intensity of the Spectra. On p. 176 it was stated that the rays from each ruling of the grating combined into a resultant ray with the phase of the ray coming from C_1 (fig. 160). The relative intensity of the spectra of the different orders depends on the variation of the amplitude of this resultant ray with θ.

If the grating is an ideal one consisting of perfectly transparent intervals of width f separated by opaque strips, then the amplitude of the resultant vibration from one interval is given by (cf. p. 172)

$$\frac{\sin \dfrac{\pi f \sin \theta}{\lambda}}{\dfrac{\pi f \sin \theta}{\lambda}}.$$

Hence when the vibrations from the N intervals are combined, the complete expression for the resultant amplitude is (cf. p. 176)

$$\frac{\sin \dfrac{\pi f \sin \theta}{\lambda}}{\dfrac{\pi f \sin \theta}{\lambda}} \times \frac{\sin \dfrac{N\pi e \sin \theta}{\lambda}}{\sin \dfrac{\pi e \sin \theta}{\lambda}}.$$

The variation of the second factor has already been discussed, and it has been shown to have maxima of equal intensity in the directions given by $e \sin \theta = n\lambda$. If we substitute these values in the first factor, it becomes

$$\frac{\sin \left(\dfrac{\pi f n}{e}\right)}{\dfrac{\pi f n}{e}}.$$

If in succession 0, 1, 2, 3 are substituted for n in this factor, and the result is squared, we obtain respectively the relative intensities of the direct image and the images of the first, second, and third orders. For $n = 0$ the expression takes the value 1. If $f = \frac{1}{2}e$, the second order spectra disappear.

If the grating has not the above ideal form, it is generally impossible to put the first factor into analytic form. If, however, the rulings are such that according to the elementary laws of reflection or refraction, apart altogether from diffraction, most of the light leaves the grating in one direction, then the spectrum formed in that direction is very bright. It is then said to be predominant. It is, for example, possible to have a grating in which half the incident light is concentrated in one of the first order spectra.

The Concave Grating. Rowland made an important step forward by ruling a grating on a spherical metal mirror, the projections on a plane of the distances between the lines being equal. Such a grating is mounted with its rulings parallel to the slit, and the spectrum is in focus without the use of lenses.

In fig. 163 AB is the grating, of course very much exaggerated in size. The radius of curvature of the grating is the diameter of the circle, and the circle touches the surface of the grating at its mid point. S is the slit. Let SA be a ray incident on the grating at A, let NA be the normal at its point of incidence, and let AP be the ray diffracted from the normal at an angle θ. Take any other point B on the grating and join SB, NB and PB; then NB is the normal to the grating. Since AB is small, B may be regarded as on the circumference of the circle. Then $\angle$NBP = $\angle$NAP since they both stand on the same arc, that is $\angle$NBP = θ. Thus no matter where B is, the rays diffracted at $\angle\theta$ pass through P, and an image of the slit is formed there without the use of lenses. The image is however an astigmatic one; a line at S at right angles to the slit is not in focus at P.

There are two ways of using the concave grating. It may be mounted in a dark room, the slit S illuminated through a hole in the wall, photographic plates set up the whole way round the arc NP and the entire spectrum photographed at one exposure. In some installations the radius of curvature of the grating is about 21 feet and the arc on which the plates are supported is an iron girder measuring as much as 29 feet round the curve. In this arrangement, when θ is large, the rays fall on the photographic plate very obliquely.

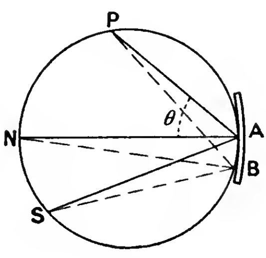


FIG. 163.

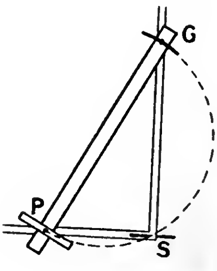


FIG. 164.

In Rowland's arrangement only one portion of the spectrum is photographed at a time. The grating G and photographic plate holder P are mounted on an iron girder facing one another, their distance apart being equal to the radius of curvature of the grating. The iron girder is pivoted at its ends on two carriages which run on rails SP and SG at right angles to one another. The slit is fixed at S and is thus on the semicircle which has PG as diameter. The rays fall on the photographic plate normally.

By moving PG along the rails different parts of the spectrum come automatically into focus at P.

In Rowland's mounting the photographic plate remains fixed at N in fig. 163 and the slit S moves along the arc between N and B.

Resolving Power of a Prism. Let fig. 165 represent the passage of a plane wave through a prism. AC is a wave-front before reaching the prism and FG a wave-front after passing through the prism.

Then since the optical distance between the wave-fronts measured along any ray is the same we have

$$AB + BF = CD + \mu DE + EG \quad . \quad . \quad . \quad . \quad (14)$$

Consider now light of a different wave-length, for which the index of refraction is $\mu + d\mu$ and draw through F the wave-front, FH, for this wave-length. Let it be inclined at a small angle $d\theta$ to FG.

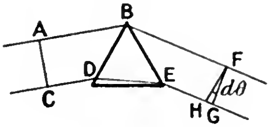


Fig. 165.

The rays BF and EH are now no longer at right angles to FH. Since, however, $d\theta$ is small, they differ in length from the rays at right angles to FH only by small quantities of the second order. We have therefore

$$AB + BF = CD + (\mu + d\mu)\, DE + EH,$$

which on subtraction from (14) gives

$$HG = d\mu\ DE, \quad \text{or} \quad a d\theta = t d\mu,$$

if we write a for FG, the width of the emergent beam, and t for DE, the thickness of prism traversed. This last equation may be written

$$\frac{d\theta}{d\mu} = \frac{t}{a} \quad . \quad . \quad . \quad . \quad . \quad (15)$$

Now consider the prism as a rectangular aperture. Each of the wave-lengths then forms a diffraction pattern of the type shown in fig. 156. If the two lines are to be resolved, the angle between them must equal the angle between the central maximum and adjacent minimum of either diffraction pattern. If we fix our attention on the first wave-length and suppose it to be diffracted in all directions from BE, the rays at right angles to FG are in the same phase on FG; consequently the principal maximum is formed in this direction. The rays at right angles to FH are not in the same phase on FH, the difference in path of the extreme rays being HG $= a d\theta$. If this is to be the direction of the first minimum, $a d\theta = \lambda$. On combining this result with (15) we obtain

$$\frac{t}{a}\, d\mu = \frac{\lambda}{a}, \quad \text{or} \quad t\,\frac{d\mu}{d\lambda}\, d\lambda = \lambda.$$

Hence the resolving power of the prism is given by

$$\frac{\lambda}{d\lambda} = t\,\frac{d\mu}{d\lambda}.$$

When the prism is used to most advantage, DE touches its base and the whole surface is filled with rays. The resolving power thus depends only on the nature of the glass and the length of the base. It is independent of the refracting angle of the prism and the lengths of the sides.

The above treatment of the resolving power of a prism is due to Rayleigh.

Resolving Power of a Telescope. As was mentioned on p. 84 when a telescope is focussed on a star, the image formed by the object glass is not a point but a small disc surrounded by rings. This is due to diffraction at the circular edge of the object glass. The theory of the phenomenon is somewhat elaborate and was first given by Airy, but its main features and rough numerical results can easily be obtained from elementary considerations. Airy's result is in the form of a series which does not lend itself readily to calculation.

Let the lens in fig. 166 represent a telescope object glass which is directed towards a star, and let the geometrical image of the star be formed at S on the axis of the instrument. Let AB be a section of the wave-front immediately after passing through the lens. It is then spherical and converging towards S. Let it cut the axis in C. Let F be the focal length of the lens and let d be its diameter. Let P be a point in the focal plane distant x from S. Consider the effect of the wave-front at P. ACB may be regarded as a straight line, and

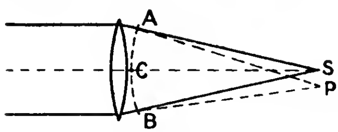


FIG. 166.

$$\mathrm{AP}^2 - \mathrm{BP}^2 = \{(\mathrm{AC} + x)^2 + \mathrm{SC}^2\} - \{(\mathrm{BC} - x)^2 + \mathrm{SC}^2\}$$
$$= 2x\mathrm{AB} \text{ approximately.}$$

Hence $$\mathrm{AP} - \mathrm{BP} = \frac{2x\mathrm{AB}}{\mathrm{AP} + \mathrm{BP}}$$
$$= \frac{xd}{\mathrm{F}} \text{ approximately.}$$

If the wave-front were cylindrical instead of spherical, all the secondary waves starting from AB would neutralise one another at P if AP – BP were equal to $n\lambda$, and the first diffraction minimum would be given by

$$x = \frac{\mathrm{F}\lambda}{d}.$$

If we were to consider only the rays in the plane of the figure and were then to rotate the figure about CS as axis, we would obtain a bright central disc surrounded by dark and bright rings, the radius of the first dark ring being given by the above formula.

The more accurate theory gives for the radius of the first dark ring

$$x = \frac{1{\cdot}22\ \mathrm{F}\lambda}{d},$$

and shows that 84 per cent of the total light of the image goes to form the central disc.

The angle subtended at the centre of the object glass by the radius of the first dark ring is $1{\cdot}22\lambda/d$ in radian measure. It is found as a result of experience, that two stars can be resolved when the centre of the disc of the one falls on the first dark ring of the image of the other. The above angle can therefore be taken as a measure of the resolving power of an object glass. Of course, if the surfaces of the lens are not sufficiently accurate and the glass not sufficiently homogeneous, a blurred image is formed instead of an accurate diffraction pattern and the resolving power is less. It depends then on the inaccuracy of the workmanship and not on the limits of the theory.

The Blue of the Sky. The blue colour of the sky is caused by diffraction. When a wave of light which is traversing space meets an obstacle such as a small water particle or a dust particle or even an oxygen molecule, the obstacle acts as a source and scatters some of the light. If the wave passes through a cloud of such obstacles, its intensity is weakened owing to loss of light by scattering. The mathematical theory of the whole process is complicated and will be omitted here. The law according to which the percentage of light scattered varies with the wave-length has, however, been derived very simply by Rayleigh in the following manner :—

Let a be the amplitude of the incident wave and let s be the amplitude of the diffracted wave at a distance of r from the obstacle. Let v be the volume of the obstacle. Then s is obviously proportional to a and inversely proportional to r ; it is plausible to assume that it is also proportional to v, since within limits a heavier obstacle will diffract more. We have therefore

$$s = \frac{kav}{r},$$

where k is a constant. Now s/a is a ratio and its dimensions are zero ; consequently the dimensions of kv/r must be zero. The dimensions of v/r are $(\text{length})^2$. Consequently the dimensions of k are $(\text{length})^{-2}$. The only quantity having the dimensions of length involved in the problem and not already considered is λ, the wave-length. Hence k varies as λ^{-2}.

The percentage of light scattered is proportional to s^2/a^2 and thus to λ^{-4}. This is almost 16 times as great for the violet end of the spectrum as for the red end. Consequently the blue contained in white light is scattered to a much greater extent than the red. The sky appears blue at a distance from the sun because the sun's rays are passing through it obliquely to our line of vision and we see it only by scattered light. On the other hand the sun and neighbouring sky appear red at sunrise and sunset because the rays coming from them have traversed a great distance in the atmosphere and consequently the blue has been diffracted to the side and lost.

The blue of the sky was imitated artificially by Tyndall. He caused light from an arc lamp to pass through a tube containing a cloud of fine particles formed by bringing nitrite of butyl vapour and hydrochloric acid vapour together at a low pressure. The size of the particles slowly increased with time. When the particles had the right size, they scattered the blue, and the path of the beam had an azure colour when viewed from the side.

The light from flames is caused by solid incandescent carbon particles of a size suitable for diffracting light and heat waves, and it has been found experimentally that when such waves are transmitted through a flame the quantity of energy lost varies as λ^{-4}.

The quantity of light scattered in any direction diminishes as the angle between this direction and the direction of vibration in the incident wave decreases. No light is scattered by the particle in a direction parallel to the direction of vibration of the incident wave.*

*The direction of vibration is here taken at right angles to the plane of polarisation, cf. next chapter. If the incident light has all possible directions of vibration, the light scattered in a particular direction at right angles to the

The distinction between regular reflection and diffraction by small particles is, that when a wave falls on a regularly reflecting surface all the secondary waves sent out by the particles on it reinforce one another and together form the regularly reflected wave. But when the wave falls on a small particle, only one secondary wave is sent out.

The halos sometimes seen surrounding the sun and moon are due to diffraction caused by small drops of water in the atmosphere.

EXAMPLES.

(1) A plane wave is passing the straight edge of a screen, the screen being at right angles to the direction of propagation of the wave. A point P is taken at a distance p from the wave-front. O, the pole of P, is distant d from the edge of the screen. Find the area of the nth zone supposing it to be cut by the edge of the screen. If p is 100 cms. find how large d must be in order that the area of successive zones may diminish by less than 1 per cent.

(2) A spherical wave of radius a has diverged from a point Q. The wave front is divided into zones with respect to a point P distant $a + b$ from Q. Show that the area of the zones is given by

$$\frac{\pi\lambda ab}{a+b}.$$

(3) A plane wave of sodium light passes perpendicularly through a circular aperture of 1 cm. diameter. Calculate the intensity and position of the diffraction bands produced in the focal plane of a microscope at a distance of 700 cms. from the aperture. (Use the graphical method suggested on p. 159. Draw the zones, move a circle representing the rim of the aperture across them, and estimate whether the area of the positive or of the negative zones predominates for each position of the circle. The amplitude may be taken as proportional to the difference in area of the positive and negative zones.)

(4) Prove that

$$\int_v^\infty \cos\tfrac{1}{2}\pi v^2 dv = \cos\tfrac{1}{2}\pi v^2\left(\frac{1}{\pi^2 v^3} - \frac{1.3.5}{\pi^4 v^7} + \frac{1.3.5.7.9}{\pi^6 v^{11}} - \dots\right)$$
$$- \sin\tfrac{1}{2}\pi v^2\left(\frac{1}{\pi v} - \frac{1.3}{\pi^3 v^5} + \frac{1.3.5.7}{\pi^5 v^9} + \dots\right).$$

(Integrate by parts.)

(5) If a system of vibrations differing in amplitude and period is superimposed, show that the resultant displacement is periodic and that its period is the least common multiple of the periods of the different vibrations.

(6) Observe the bands produced by diffraction at a narrow slit using the arrangement described on p. 173. Measure their angular separation and compare it with the theoretical value. (A sodium flame will probably not be bright enough; use an incandescent mantle with a sheet of red glass in front of it.)

(7) A very large opaque screen contains a small rectangular opening. Parallel monochromatic light, incident normally on the screen, passes through the aperture. Investigate the diffraction phenomena produced.

(8) The slit of a spectrometer is illuminated with sodium light and a transmission diffraction grating placed upon the prism table with its surface

direction of incidence always vibrates at right angles to the plane containing the direction of incidence and direction of scattering. Hence the light received from the sky is polarised.

at right angles to the axis of the collimator. The distance between two adjacent rulings is e. The angle between the direct image and the first image is θ. The slit is then narrowed and the D lines seen separate. Their angular separation is measured and found to be $d\theta$ in radian measure. Show that their difference of wave-length is given by $e \cos \theta \, d\theta$.

(9) The slit of a spectrometer is illuminated with sodium light and a transmission grating placed somewhat inaccurately upon the prism table, so that the normal to its surface makes a small angle ϕ with the axis of the collimator. The angular separation between the direct image and the first image is found in the usual way to be θ. Show that

$$\lambda = e \sin \theta \pm e\phi(1 - \cos \theta),$$

the sign changing with the side on which the first image is taken. Hence show that if the determination is made first from the image on the one side and then from the image on the other, and if the mean of the two results is then taken, the error due to the inaccurate setting cuts out.

Verify the formula experimentally.

(10) Mount a Thorp diffraction grating on a spectrometer table and move a piece of cardboard across the front of its surface until the D lines are just seen separate (*a*) in the first order spectrum, (*b*) in the second order spectrum. Calculate the theoretical resolving power in each case from the fraction of the grating covered, and see how it compares with the experimental value.

(11) Repeat the previous exercise with a prism instead of a grating.

(12) Sometimes, in order that the spectrum may undergo no deviation, a replica of a grating is mounted on the face of a glass prism with its rulings parallel to the refracting edge. Show how to calculate the angle of the prism in order that any particular line in the spectrum may be undeviated. Assume that the grating is on the second face of the prism and that the light falls on the first face normally.

(13) The index of refraction of a certain glass is given by

$$\mu_C = 1{\cdot}6545, \quad \mu_D = 1{\cdot}6585, \quad \mu_E = 1{\cdot}6635.$$

The wave-lengths of the C and E lines in the solar spectrum are respectively 6563 and 5270 10^{-8} cms. Calculate the length of the base of a 60° prism made of this glass, which is just capable of resolving the D lines. The wave-lengths of the D lines are 5890 and 5896 10^{-8} cms.

(14) A 60° prism is to be made of theoretical resolving power just sufficient to separate the D lines. Calculate the length of its base provided that the refracting material is to be (*a*) crown glass, (*b*) flint glass, (*c*) carbon bisulphide.

CHAPTER XI.

POLARISATION AND DOUBLE REFRACTION.

In 1669 Erasmus Bartholinus discovered, that when a ray of light is refracted by a crystal of calcite it forms two refracted rays. To this phenomenon the name of double refraction has been given. It is exhibited by many other substances besides calcite, but historically calcite was the point of departure of the development of the whole subject, and in calcite double refraction is very marked and easy to study experimentally. We shall commence therefore by describing the phenomena shown by calcite.

Calcite or Iceland spar is crystallised calcium carbonate and was at one time found in great quantities in Iceland in very large crystals of watery clearness. It cleaves very perfectly along three directions forming parallelipipeds, as they would be termed in geometry, with their faces parallel to the planes of cleavage; these parallelipipeds are in this connection always called rhombohedra or rhombs. The angles of the parallelograms forming the sides of the rhombs are 102° and 78°, more accurately 101° 53′ and 78° 7′. At two opposite corners of the rhomb three angles of 102° meet; at the other corners one angle of 102° and two angles of 78° meet. The relative lengths of the edges of the rhomb are immaterial and may have any values. The optic axis is a direction in the crystal parallel to the straight line through a blunt corner of the rhomb, which makes equal angles with the three edges meeting there. Thus an optic axis can be drawn through every point in the crystal. A principal section is a plane through the optic axis perpendicular to two opposite faces of the rhomb. A rhomb has thus three principal sections through every point. If two opposite faces have equal sides, the principal section at right angles to these faces is parallel to the short diagonals of these faces. The above relations and definitions cannot be made clear by a diagram on the flat; consequently every student should make a model of a rhomb for himself, either in cardboard or wood or soap.

If we have an illuminated distant aperture and we look at it through a calcite rhomb, keeping the faces that we look through at right angles to the direction of the aperture, we see two images of the aperture. The line joining them is in a principal section of the rhomb. Fig. 167 represents the face of the rhomb next the eye; for the sake of simplicity it is drawn with equal sides. O and E are the two images;

they are shown in the shorter diagonal of the face, though it is only necessary that the line joining them should be parallel to the latter. If the rhomb is rotated round the direction of the distant aperture as axis, the image O remains stationary and the image E moves round it in a circle keeping always in the shorter diagonal of the end face.

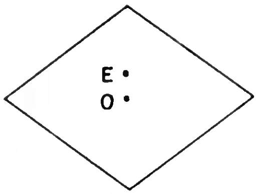


FIG. 167.

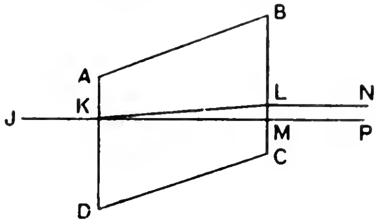


FIG. 168.

Let us consider this from another standpoint. Let ABCD be a section of the rhomb by the principal plane in which O and E lie. The distant aperture is in the same plane. Let JK be the pencil of rays from it incident normally on AD at K. The rays KP forming the image O are not displaced by their passage through the rhomb, since that image is stationary when the figure is rotated round JP as axis. The rays forming the image E must follow the path KLN, since that image moves in a circle round JP when the figure is rotated round JP as axis. LN is, of course, parallel to MP.

The ray KM that obeys the ordinary laws of refraction is called the ordinary ray and the image it forms the ordinary image; the ray KL is called the extraordinary ray and the image it forms the extraordinary image.

§ Suppose now, that instead of one rhomb the distant aperture is regarded through two rhombs the surfaces of which are parallel, as is represented in fig. 169. Then the ordinary ray traverses the second

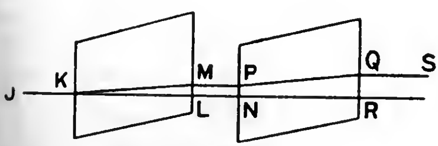


FIG. 169.

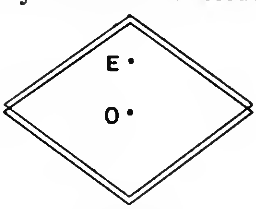


FIG. 170.

rhomb undeviated, but the path of the extraordinary ray inside the second rhomb is parallel to its path inside the first, so that, if the rhombs are equally thick, as we shall suppose for the sake of simplicity, after passing through the second rhomb the displacement between the rays has been doubled. Fig. 170 represents this result diagrammatically; it shows the end sections of the two rhombs parallel and OE, the distance between the images, twice as great as in fig. 167.

If, now, the first rhomb is kept fixed but the rhomb nearer the eye rotated through a small angle about the undeviated ray as axis, instead of two images we obtain four (fig. 171). The image O remains undeviated, the image E moves to the side, O and E both become slightly fainter, and between O and E appear two faint new images O_1

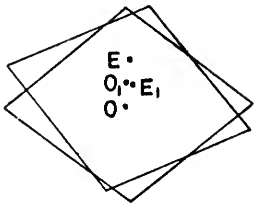


Fig. 171.

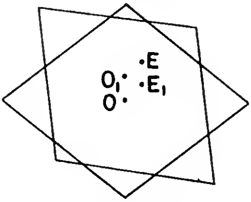


Fig. 172.

and E_1. OO_1EE_1 forms a parallelogram, the sides of which remain parallel to the principal sections of the two rhombs.

In fig. 172 the second rhomb has been rotated round so that its principal plane makes 45° with the principal plane of the first. The four images are now equally bright. If the rotation is continued

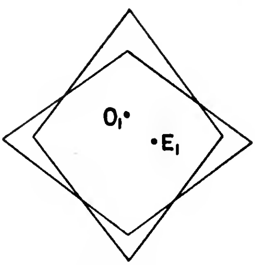


Fig. 173.

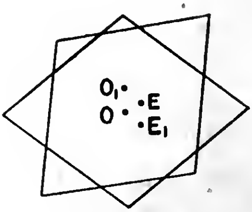


Fig. 174.

further, O and E decrease in intensity and O_1 and E_1 increase in intensity, until finally at 90° (fig. 173) O and E disappear altogether.

After passing through 90° O and E reappear and increase in intensity while O_1 and E_1 begin to decrease in intensity. Fig. 174 shows the appearance at 135° when the four images are once more equally bright. Fig. 175 shows the appearance at 180°; the two principal sections are once more parallel, the images O_1 and E_1 have disappeared and the images O and E have once more attained their initial intensities but are superimposed. As the rotation is continued from 180° to 360° the same changes are gone through in the reverse order.

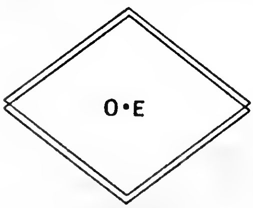


Fig. 175.

From the displacements it is obvious that the ordinary image formed by the first rhomb has been decomposed by its passage through the second into an ordinary O and an extraordinary E_1, whilst the extraordinary image formed by the first rhomb has been decomposed by its passage through the second into an ordinary O_1 and an extraordinary E. The question now arises as to what causes the variation in intensity.

§ Hitherto in the chapters on interference and diffraction it has been necessary to assume only that light was propagated by wave-motion without specifying the direction of the vibrations. Huygens and Young thought that they were longitudinal, that they took place in the direction in which the wave travelled. Fresnel first showed that they were transverse and thus completely explained the variation of intensity in the experiment with the two rhombs.

Let us assume that the ordinary ray emerging from the first calcite rhomb consists of rectilinear vibrations parallel to the longer diagonal of the end face, that the extraordinary ray consists of rectilinear vibrations parallel to the shorter diagonal, and that these vibrations are of equal intensity. Then their amplitudes may be represented by two equal straight lines, PO and PE, of length a at right angles to one another. Suppose now that the principal plane of the second rhomb makes $\angle\theta$ with the principal plane of the first. The vibration PO decomposes into one PN of amplitude $a \cos \theta$ at right angles to the principal plane of the second rhomb and one NO of amplitude $a \sin \theta$ parallel to this direction. It is these vibrations that give respectively the images O and E_1. The vibration PE decomposes into one PM parallel to the principal plane of the second rhomb, of amplitude $a \cos \theta$ and one ME perpendicular to it, of amplitude $a \sin \theta$. These vibrations give respectively the images E and O_1. Thus the intensities of O and E are each $a^2 \cos^2 \theta$, and so these images vanish when the rhombs are crossed, i.e. when $\theta = 90°$, while the intensities of O_1 and E_1 are each $a^2 \sin^2 \theta$, and consequently these images vanish when the rhombs are parallel. The variation of the intensities in the intermediate positions also agrees with the above formulæ.

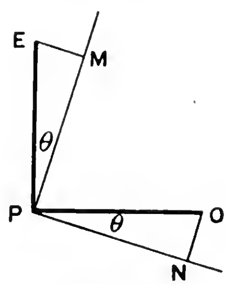


Fig. 176.

It is obvious, however, that we could have the same agreement by making the opposite assumption, namely, that the ordinary ray vibrated in the principal plane and the extraordinary ray vibrated at right angles to the principal plane. The one explanation has no advantage over the other. We therefore include them both in the statement, that the ordinary ray is plane polarised in the principal plane and that the extraordinary ray is plane polarised at right angles to the principal plane. This leaves the question open as to which of the two planes the vibrations take place in respectively. Over this point there was a long difference of opinion.

In the expression taken to represent a light wave on p. 131 the direction of η was left undefined. If we take it at right angles to the axis, the expression represents a plane polarised wave.

The experiment with the two rhombs may be regarded then as proof that the ordinary and extraordinary rays vibrate in directions at right angles to one another and to the direction of propagation of the ray. This conclusion was very thoroughly tested and verified by Fresnel and Arago in a celebrated series of experiments on interference. They found that:—

(1) Two rays of light polarised at right angles do not interfere under the same conditions as two rays of ordinary light.

(2) Two rays of light polarised in the same plane interfere like two rays of ordinary light.

(3) Two rays polarised at right angles may be brought to the same plane of polarisation without thereby acquiring the property of being able to interfere with each other.

(4) Two rays polarised at right angles and afterwards brought to the same plane of polarisation interfere like ordinary light if they originally belonged to the same beam of polarised light.

Natural Light. The question arises now as to the constitution of unpolarised natural light, that is, of the light before it falls on the calcite rhomb. The fact that the ordinary and extraordinary images formed by a calcite rhomb are always equally bright, shows that the incident light possesses no one-sidedness. It is simplest, therefore, to regard it as consisting of plane polarised light, the direction of the plane of polarisation of which undergoes sudden and irregular changes. If these changes occurred seldom, once every five or ten seconds or thereabouts, and we watched the two images formed by the rhomb, their intensities would be unequal, that one being brighter the plane of polarisation of which formed the smaller angle with the plane of polarisation of the incident light. Also, whenever the direction of the plane of polarisation of the incident light changed, the relative intensity of the two images would change. If, however, the changes take place too frequently for the eye to follow them, and if in the course of $\frac{1}{20}$ second the direction of the plane of polarisation of the incident light undergoes so many random changes, that on the average it has every possible direction for the same interval of time, then the eye can detect no difference in the intensities of the ordinary and extraordinary images. Such, therefore, must be the nature of unpolarised light.

Means of Producing Plane Polarised Light. The separation of the ordinary and extraordinary beams is much greater for Iceland spar than for most crystals, and an Iceland spar rhomb may be used as a means of producing plane polarised light simply by placing a screen to absorb the one beam on its emergence from the prism. This method is not practicable unless the rhomb is a large one and the beam a narrow one.

In certain kinds of tourmaline, a double refracting crystal which occurs in many different colours, the ordinary ray is so strongly absorbed that plates cut parallel to the axis practically transmit no light except the extraordinary. Such plates can therefore be used as a means of producing plane polarised light.

In 1808 Malus discovered, that when light is reflected at a particular angle from the surface of glass, water, or other transparent substances, it is almost completely plane polarised in the plane of incidence. This particular angle of incidence is called the polarising angle. A few years later Brewster showed that the tangent of the polarising angle was equal to the index of refraction of the medium in question. This fact, Brewster's law as it is called, leads to the result that at the polarising angle the reflected and refracted rays are at right angles to one another. For, let ϕ be the angle of incidence and θ the angle of refraction; then

$$\frac{\sin\phi}{\sin\theta} = \mu.$$

But by Brewster's law $\mu = \tan\phi$; therefore

$$\frac{\sin\phi}{\sin\theta} = \tan\phi,\ \sin\theta = \cos\phi,$$

and $\phi + \theta = \frac{\pi}{2}$, which gives the required result.

The polarisation of light reflected at the polarising angle is never quite perfect, there being always some natural light mixed with it, especially in the case of substances of high refractive index. This has been shown to be due to the formation of a film of lower refractive index on the surface, either by polishing or by weathering of the surface or by dirt.

§ Let us now return to the propagation of light in calcite. Take any plane through the axis and cut on the calcite a plane face at right angles to it. We shall now extend the definition of principal section to include such a plane through the axis perpendicular to any face cut on the crystal; as previously defined it had to be perpendicular to a cleavage face. Calcite is thus a crystal having an infinite number of principal sections all intersecting in the optic axis.

If, now, a ray of light is incident in the principal section on any plane face cut in a calcite crystal, and if the angle of incidence is varied, it is found that the incident ray gives rise in general to two refracted rays, and one of these, the ordinary ray, obeys the ordinary laws of refraction. The ratio of the sine of the angle of incidence to the sine of its angle of refraction has a constant value, μ_o, which in the case of sodium light is equal to 1·6584. The other ray, the extraordinary ray, lies in the principal section, but the ratio of the sine of the angle of incidence to the sine of its angle of refraction is not constant.

If from a point C (fig. 177) within the crystal ordinary rays are propagated in a principal section in all directions the wave-front is a circle BB_1. Huygens demonstrated experimentally that, if from the same point extraordinary rays are propagated in the same plane in all directions, their velocity varies as the radius-vector of an ellipse, being least and equal to the velocity of the ordinary ray in the direction of the axis and greatest at right angles to this direction. The extraordinary wave-front is consequently an ellipse touching the circle in the optic axis BB_1 and having the radius of the circle as its semi-axis minor. The ratio $AC/CB = \mu_o/\mu_e$, where μ_e is a constant which has the value 1·4864 for sodium light.

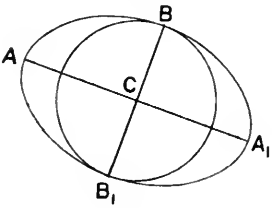


FIG. 177.

Fig. 177 holds for all planes through BB_1 since they are all principal sections. If the figure is rotated about BB_1, it gives a sphere and an ellipsoid of rotation. Hence, if a light motion begins at any point inside an Iceland spar crystal and advances unhindered in all directions, the wave-front is a double surface consisting of a sphere and an ellipsoid of rotation. The latter has the optic axis as minor axis and as axis of rotation, and it touches the sphere at the ends of its minor axis.

§ Let us consider the propagation of the extraordinary wave a little more closely. Assume that it emanates from the point C but that the source is active only for a very short interval. Then after a time the disturbance is confined between two concentric similar ellipsoids. The solid figure bounded by these ellipsoids is called an ellipsoidal shell. The rays are in general not perpendicular to the wave-front; only at the extremities of the axes are the radius-vectors of an ellipse perpendicular to the curve. Let PN be the normal at the point P, and at this point let a screen with an aperture be placed. A portion of the wave-front passes through the aperture and is shown in successive positions at D and E. It is moving away from the normal at P. The energy travels along the ray and not along the normal to the wave-front.

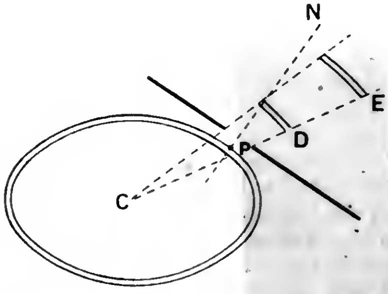


FIG. 178.

Let A, B, and C be three points on a plane wave-front in a calcite crystal. Then by Huygens' principle these points may be regarded as secondary sources sending out ellipsoidal waves. The optic axis through each point is at right angles to the dotted line. The envelope

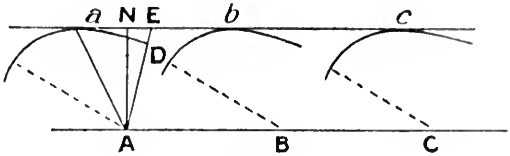


FIG. 179.

of the secondary waves is obviously the plane represented in the diagram by *a*, *b*, *c*. A*a*, B*b*, and C*c* are rays from the one wave-front to the other. The rays are the shortest optical distances between the two wave-fronts. For, consider any other direction AE; AE is greater than AD and AD is optically equivalent to A*a*, hence AE is optically greater than A*a*.

The velocity with which the wave-front moves forward in the direction of the normal is called the wave velocity. It is obvious from the figure that the ratio of the wave velocity to the ray velocity is given by AN/A*a*.

It has been assumed for simplicity in fig. 179 that the optic axis is in the plane of the paper. This is of course not necessary. In the general case the plane containing the rays A*a*, B*b*, and C*c*, is not perpendicular to the wave-front, e.g. the point *a* is not only displaced to the side of N but also in front of it or behind it.

Refraction at the Surface of Calcite. If a plane wave is incident from air on a plane face of a calcite crystal cut in any direction whatever with respect to the axis, the two wave-fronts in the crystal can be determined by Huygens' principle, in perfect analogy with the construction for isotropic media given on p. 128.

For let CB (fig. 180) be the incident wave-front, and let CP be the intersection of the surface with the plane of incidence. Draw BP at right angles to CB. The direction of the optic axis at C is known. We shall suppose it to be in the plane of incidence and represent it by the dotted line. Regard C as a secondary source; with it as centre describe a sphere with radius PB/μ_o and also an ellipsoid of revolution, the semi-axes of which are given by PB/μ_o and PB/μ_e and which touches the sphere at its intersections with the dotted line and has the dotted line as axis of revolution. Then at the instant when the disturbance from B has reached P, the disturbance from C has reached the surfaces of the sphere and ellipsoid.

In the same way spheres and ellipsoids can be drawn for points on CP between C and P, and it is found that these spheres and ellipsoids have common tangent planes through P. If therefore we draw tangent planes through P to the secondary wave-fronts from C, these tangent planes are the two refracted wave-fronts. The lines CO and CE to the points of contact are respectively the ordinary and extraordinary refracted rays, that correspond to the incident ray through C.

In the case of the ordinary wave the sine of the angle of incidence is to the sine of the angle of refraction as the velocity in air to the velocity of the ordinary wave in the crystal. Similarly, if a normal is drawn through C to the extraordinary wave-front, the sine of the

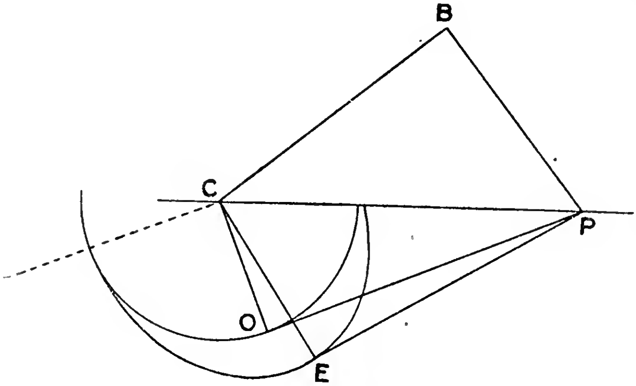


FIG. 180.

angle of incidence is to the sine of the angle of refraction of this normal as the velocity in air is to the wave velocity in the crystal. But the ratio of the sine of the angle of incidence to the sine of the angle of refraction of the extraordinary ray is not the same as the ratio of the velocity in air to the velocity of the refracted extraordinary ray. This can easily be shown by drawing a straight line through E at right angles to CE to meet CP in Q.

If the optic axis is not in the plane of incidence, that is, if the latter is not a principal section, the tangent plane does not in general touch the ellipsoid in the plane of incidence. The extraordinary refracted ray is therefore not in the plane of incidence.

If the optic axis and incident ray are both perpendicular to the surface, the ordinary and extraordinary rays have obviously the same direction and velocity. In this case the light is propagated inside the crystal as natural light. If a parallel-sided plate of calcite is cut with its faces normal to the axis and a beam of light is incident on it perpendicularly, then it goes through it as if it were a plate of glass.

Optic Axis at Right Angles to the Plane of Incidence. If the optic axis is at right angles to the plane of incidence (fig. 181) the latter is a diametral plane of the ellipsoid and intersects it in a circle. Consequently the extraordinary refracted ray lies in the plane of incidence, is normal to its wave-front and behaves as if it had a constant index of refraction μ_e. The latter quantity is hence called the extraordinary

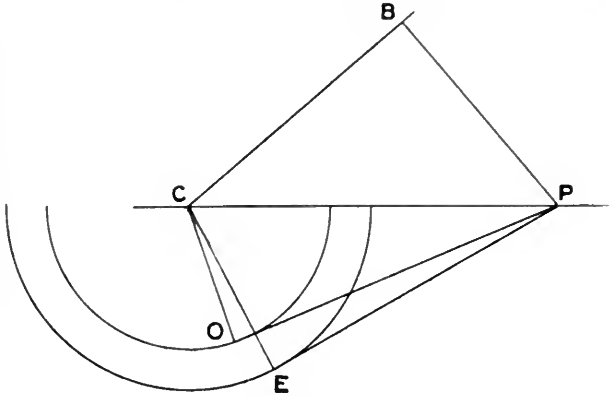


FIG. 181.

index of refraction of the crystal; μ_o and μ_e are called the two principal indices of refraction.

Optic Axis in the Surface and Plane of Incidence. In this case (fig. 182) the refracted rays are both in the plane of incidence. Since

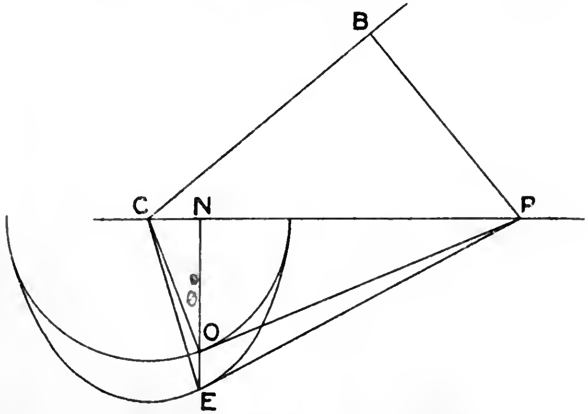


FIG. 182.

the ellipse can be projected into the circle and at the same time the tangent PE projects into the tangent PO, EO produced meets CP at right angles in N and NE/NO $= \mu_o/\mu_e$. If θ_o and θ_e are respectively the angles of refraction of the ordinary and extraordinary rays, $\theta_o = \angle$CON, $\theta_e = \angle$CEN, and it follows that

$$\frac{\tan\theta_o}{\tan\theta_e} = \frac{CN/NO}{CN/NE} = \frac{NE}{NO} = \frac{\mu_o}{\mu_e}.$$

Experimental Determination of μ_o and μ_e. If a calcite prism is cut either with the optic axis parallel to the refracting edge or with the optic axis bisecting the refracting angle, then rays traversing the prism at minimum deviation must be perpendicular to the optic axis and consequently must be refracted like ordinary light with indices of refraction μ_o and μ_e. If such a prism is placed upon the spectrometer table and the slit illuminated with monochromatic light, two images of the slit are seen, and if each is adjusted in succession for minimum deviation and the calculation made exactly as in the same way as for a glass prism, μ_o and μ_e are obtained.

The Nicol Prism. The Nicol prism, called after its inventor, is the most commonly used means of producing plane polarised light that we have. It consists of a natural rhomb of Iceland spar, the edges of the end faces of which are equal and one-third the length of the other edges. It is sliced from the one blunt corner to the other in a plane parallel to the long diagonal of the end faces; the cut faces are polished and re-united with a film of Canada balsam. Fig. 183 represents the

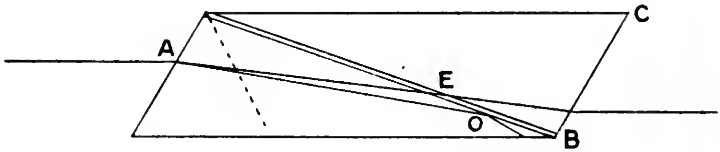


FIG. 183.

arrangement. The dotted line gives the optic axis. If a ray of light is incident at A in the plane of the figure, it is decomposed into an ordinary ray AO and an extraordinary ray AE. The mean refractive index of the ordinary ray is 1·66 and the mean refractive index of Canada balsam is 1·54. The rays are in general incident on the film at such an angle that the ordinary ray is totally reflected to the side and the extraordinary transmitted. The side of the rhomb is blackened and the ordinary ray absorbed there. Consequently the transmitted light is polarised in a plane at right angles to the principal section of the rhomb.

The beam of light entering a nicol is usually convergent. Consequently all the rays are not incident on the film at the same angle, and if the angle of convergence is too great, some of the extraordinary rays that should emerge near B are totally reflected and some ordinary rays emerge near C. The greatest admissible angle between the extreme rays of the incident beam, measured in air, is 24°.

As the obliquity of the end faces causes a displacement of the image when the nicol is rotated, it is now customary to cut the ends perpendicular to the long sides. Other modifications have also been made; for a description of these an article by S. P. Thompson ("Proceedings of the Optical Convention," 1905) should be consulted.

Double Image Prisms. Fig. 184 represents a Rochon's prism which consists of two equal prisms of calcite or other doubly refracting substance, cut differently with respect to the axis. The incident ray AB is normal to the surface. The optic axis in the first prism, as shown in the figure, is parallel to the incident light; consequently in that prism both rays are transmitted with the same velocity. In the second prism the optic axis, as shown in the figure, is at right angles to the plane of the paper; the ordinary ray is therefore transmitted undeviated but the extraordinary ray is deflected to the side.

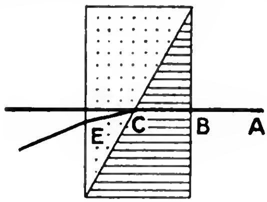


FIG. 184.

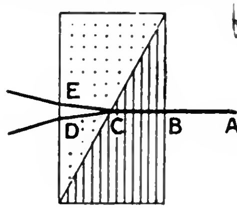


FIG. 185.

In Wollaston's prism the optic axis of the first prism is in the plane of incidence and perpendicular to the incident ray. Consequently, although the ordinary and extraordinary rays have the same direction in it, their velocities are different. In the second prism the optic axis is perpendicular to the plane of the figure. On entering it the ordinary ray becomes the extraordinary; and it is refracted in the direction CD, while the extraordinary ray becomes the ordinary, and it is refracted in the direction CE.

Other Uniaxal Crystals. Since Huygens' time it has been shown for other crystals besides calcite that the wave surface consists of a sphere and an ellipsoid of revolution touching the sphere at the ends of its axis of revolution. In all these crystals, as in calcite, there is only one direction without double refraction, and hence they are said to be optically uniaxal crystals.

Optically uniaxal crystals are divided into two classes, those like calcite, for which μ_o is greater than μ_e, and those like quartz, for which μ_o is less than μ_e. For the first class the ellipsoid is outside the sphere and is an oblate spheroid; for the second the ellipsoid is inside the sphere and is a prolate spheroid. The crystals of the first class are said to be negative because, as may be seen by graphical construction, in them the extraordinary ray makes a greater angle with the axis than the ordinary ray does. It is thus "repelled" by the axis. The crystals of the second class are called positive because in them the extraordinary ray is between the ordinary ray and the axis; it is "attracted" by the axis.

In all optically uniaxal crystals the ordinary ray is polarised in the principal section, and although the values of μ_o and μ_e vary slightly with the wave-length, the direction of the optic axis is independent of the wave-length.

§ Let us assume that two nicols are placed one after the other on the same axis and that a source of light is regarded through them. Then its apparent intensity depends on the orientation of the principal sections of the two nicols to one another.

Let OP (fig. 186) denote the principal section of the first nicol and OA the principal section of the second nicol, and let $\angle AOP = \theta$. Let a be the amplitude of the light after emerging from the first nicol. Then, if reflection and absorption losses in the second nicol are disregarded, the amplitude after emerging from the latter is $a \cos \theta$. If reflection and absorption losses in the first nicol are disregarded, the amplitude of the natural light incident on it may be taken as $a \sqrt{2}$. Thus the intensity of the incident beam is $2a^2$ and the intensity of the beam after emerging from the second nicol is $a^2 \cos^2 \theta$.

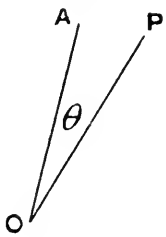


FIG. 186.

If the second nicol is rotated about its axis, that is, about the direction of the source of light, the brightness of the latter becomes a maximum twice in the course of a complete revolution when $\theta = 0$ or π. The nicols are then said to be parallel. The field becomes wholly dark twice in a complete revolution, when $\theta = \frac{\pi}{2}$ or $\frac{3\pi}{2}$. The nicols are then said to be crossed.

If a glass plate is placed between the two nicols, the experiment is not affected, but if a plate of a doubly refracting substance is interposed, the image is in general not extinguished when the nicols are crossed. Examination between crossed nicols is thus a means of detecting the presence of double refraction. In this connection the first nicol, the one whose principal section is denoted by OP, is called the polarising nicol or polariser and the second the analysing nicol or analyser.

Interference in Crystal Plates in Polarised Light. We shall now consider the appearance in the field when a plate with plane parallel faces cut from a uniaxal crystal is placed between the nicols. To simplify the calculation the following assumptions are made:—

(1) The light reflected at the surfaces of the plate and nicols is disregarded.

(2) The ordinary and extraordinary rays have the same path inside the crystal although their velocities are different.

If LH represents the incident ray and HKT and HSV the ordinary and extraordinary rays, after emerging from the plate the latter are parallel and polarised in planes at right angles to one another. According to the second assumption KS is so small that we take S to coincide with K.

Let v_1 and v_2 be the velocities of the ordinary and extraordinary rays in the crystal, let v be the velocity of light in air, and let d = HK = HS.

Write $v/v_1 = \mu_1$ and $v/v_2 = \mu_2$. Then the phase difference of the two rays after passing through the plate is given by

$$\delta = \frac{2\pi}{\tau}\left(\frac{d}{v_1} - \frac{d}{v_2}\right) = \frac{2\pi d}{\lambda}(\mu_1 - \mu_2) \quad . \quad . \quad (16)$$

where λ is the wave-length of the light in air.

Let the amplitude of the light before it falls on the plate be unity and let OP be its plane of polarisation. Let OX and OY be respectively the planes of polarisation of the ordinary and extraordinary

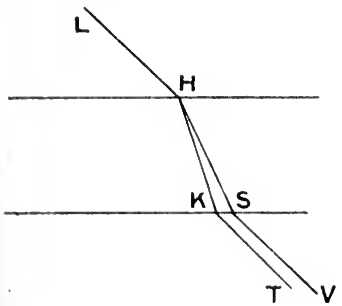


FIG. 187.

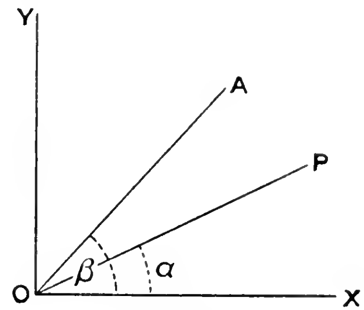


FIG. 188.

rays after emerging from the plate. Then their amplitudes are $\cos\alpha$ and $\sin\alpha$, if $\angle$XOP be denoted by α. During their passage through the plate they have acquired a relative phase difference δ. Let OA be the plane of polarisation of the analysing nicol and let $\angle$XOA be denoted by β. Only the components of the ordinary and extraordinary rays that are polarised in the plane OA are transmitted by the second nicol. They have amplitudes $\cos\alpha\cos\beta$ and $\sin\alpha\sin\beta$ and a relative phase difference δ; consequently they compound into a single ray with the same period and an intensity I given by

$$\begin{aligned} I &= \cos^2\alpha\cos^2\beta + \sin^2\alpha\sin^2\beta + 2\cos\alpha\sin\alpha\cos\beta\sin\beta\cos\delta \\ &= (\cos\alpha\cos\beta + \sin\alpha\sin\beta)^2 - 2\cos\alpha\sin\alpha\cos\beta\sin\beta(1 - \cos\delta) \\ &= \cos^2(\alpha - \beta) - \sin 2\alpha\sin 2\beta\sin^2\frac{\delta}{2} \quad . \quad . \quad . \quad . \quad . \quad (17) \end{aligned}$$

If the nicols are parallel, $\alpha = \beta$ and

$$I = 1 - \sin^2 2\alpha\sin^2\frac{\delta}{2}, \quad . \quad . \quad . \quad (18)$$

If the nicols are crossed, $\alpha = \beta \pm \frac{\pi}{2}$ and

$$I = \sin^2 2\alpha\sin^2\frac{\delta}{2} \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Thus the intensities in these two positions are complementary.

We shall now make applications of the above formulæ, but it will make things clearer if before doing so a piece of apparatus * commonly used is described. It is represented in fig. 189, is more convenient

* Usually known as Norremberg's apparatus.

than two nicols and possesses the virtues of simplicity and cheapness, though it is not the most refined piece of apparatus for investigating doubly refracting substances in polarised light. N is a nicol which rotates about a vertical axis and the position of which can be read on a horizontal circle. A is an aperture in a horizontal plate. P is a glass plate that can be rotated about a horizontal axis and the position of which can be read on the circle C. L is a lens which can be swung into the position L′ in the path of the rays, and M is a piece of mirror glass.

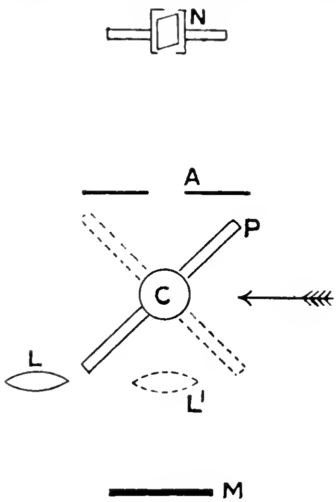


Fig. 189.

It is usual to describe the application of the formulæ (17), (18), and (19) under two heads according as the crystal plate is traversed by a parallel or a convergent beam of light.

It has already been mentioned, that plates of tourmaline cut parallel to the axis absorb the ordinary ray so strongly that only the extraordinary ray is transmitted. Hence two plates of tourmaline held in wire supports and known as tourmaline tongs or forceps are often used for observing the interference figures produced by crystals. The crystal is placed between the tourmalines, both of which can be rotated separately, the combination is held close to the eye and the observer looks through it at a white cloud or other uniformly illuminated extended object.

Parallel Light. In this case the apparatus can be used in two ways. Light is incident in the direction of the arrow either from a sodium flame or from the sky. If a sodium flame is used, it is well to put a ground glass plate in front of it, as the extent of the flame is not usually great enough. The glass plate P is set in either of the two positions shown at such an angle that the incident light is plane polarised by reflection. If the glass plate is in the position shown by the full lines, the light then falls on the mirror—the lens is swung out to the side—is reflected there and passes up through the plate P and the aperture A to the nicol and through it to the eye of the observer. The crystal plate is placed on the mirror M and consequently the rays traverse it twice, once on the downward and once on the upward path.

If the glass plate is in the position shown by the dotted lines the light after reflection passes directly through the aperture to the nicol. In this case the crystal plate is placed over A and the light traverses it once.

The glass plate P takes the place of the polarising nicol. After reflection from it the beam is plane polarised in the plane of the diagram.

On entering the crystal the beam decomposes into two, plane polarised at right angles to one another, which travel with different velocities. Consequently after emerging from the crystal they have acquired a relative phase difference, but as they are polarised in different planes they are not in a condition to interfere until the analysing nicol brings them into the same plane and makes interference possible.

The eye is focussed on the crystal plate itself, and as the pencils of light coming from the different points on its surface are thin and the plate subtends a small angle at the eye, the rays that reach the eye practically all pass through the plate normally. Hence δ is the same for every point on its surface, and if the illumination is monochromatic, the plate appears of a uniform brightness. If the incident light is white, owing to δ varying with the wave-length the different components of the white light are not all transmitted to the same extent, and the plate appears coloured. The colour depends on the thickness and is purest for the case of " crossed nicols ".

Quartz Wedge. The variation of the colour with the thickness can be studied when a wedge is used instead of a plate. Quartz wedges are commonly used for this purpose ; the angle between their faces is very small, usually about $\frac{1}{2}°$, and to prevent breaking they are cemented to glass plates. Owing to the angle between the faces being so small we can assume that it does not affect the directions of the reflected or transmitted light appreciably but only the value of δ ; the latter increases from the thin end to the thick end of the wedge. The wedge is cut with the optic axis parallel to its sides. μ_1 and μ_2 in (16) then become the principal indices of refraction for quartz. Their values for sodium light are 1·5442 for the ordinary ray and 1·5533 for the extraordinary ray ; the difference is ·0091. It increases slightly towards the violet, being ·0094 for that end of the spectrum.

Let us suppose that the nicol is crossed and that the wedge is set with its axis at 45° to its principal plane. Then $\sin^2 2\alpha = 1$, and the expression for the intensity becomes simply $I = \sin^2 \frac{\delta}{2}$ and is a maximum when $\delta = (2n + 1)\pi$, i.e. $d = \frac{(n + \frac{1}{2})\lambda}{\mu_1 - \mu_2}$,

or
$$d = \frac{(n + \frac{1}{2})\lambda}{\cdot 0091} \qquad . \quad . \quad (20)$$

if the variation of μ with λ is neglected. The different colours have their maxima at different places and the wedge is crossed with coloured bands parallel to its edge. If at the thin end d is less than $\frac{1}{2}\lambda/\cdot 0091$ where λ has its value for the violet, then the thin end is dark. It will be remembered that when Newton's rings were seen in white light (p. 143) the maxima of the different colours were given by

$$2e = (n + \tfrac{1}{2})\lambda$$

where e was the thickness of the air film. This equation is of the same type as (20). Thus the sequence of the colours exhibited by the quartz wedge is somewhat the same as that shown by Newton's rings. Hence if $\mu_1 - \mu_2$ is known, the thickness can be determined from the colour, and, if the thickness is known, $\mu_1 - \mu_2$ or the degree of birefringence can be determined from the colour.

If the illumination is monochromatic, the wedge is simply crossed by equally spaced bright and dark bands. If the analysing nicol is turned through 90°, the positions of the bright and dark bands are interchanged. If the nicol is kept fixed and the wedge rotated through one complete revolution, i.e. if we increase α and β each 2π, keeping their difference constant, the bands disappear eight times in the revolution, i.e. whenever $\sin 2\alpha$ or $\sin 2\beta$ becomes zero.

Convergent Light. To use the apparatus shown in fig. 189 with convergent light the crystal plate is placed at A, the nicol pushed down close to A and the glass plate turned into the dotted position to receive light from the sky. The eye looks through N and the crystal at different points in the sky. It is focussed on the sky.

We shall suppose that the plate is cut at right angles to the optic axis and that E (fig. 190) represents the position of the eye. Then

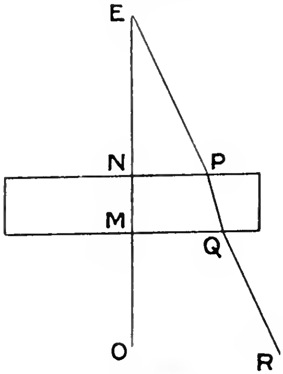


FIG. 190.

FIG. 191.

for the rays passing normally through the plate in the direction OMNE $\delta = 0$. For rays passing through the plate in the direction RQPE δ is constant as long as ∠NEP is constant, because the velocities in the crystal depend only on the inclination of the path to the optic axis. δ increases with the inclination of the ray to the axis. Hence, as the colour depends on δ, the eye sees the direction EN surrounded by coloured circles. The lines of an interference figure which exhibit the same colour at all points are called isochromatic curves.

If the nicol is crossed, the intensity is zero when $\sin^2 2\alpha = 0$. Hence superimposed on this system of circles is a dark cross. The arms of this cross are referred to as achromatic lines. If the nicol is parallel we have, of course, the complementary figure and the achromatic lines are displaced through 45°. Fig. 191 shows the interference figure when the light is monochromatic.

A crystal plate can also be examined in convergent light when placed on the mirror M (fig. 189) if the lens is swung into the path of the rays and focussed on M.

The apparatus represented in fig. 189 is, of course, suitable only for large crystals. For small crystals a microscope must be used; for this purpose it is provided with a polarising nicol below the stage and an analysing nicol either immediately above the objective or above the eyepiece. Then, for examination in parallel light, the microscope is focussed on the crystal. For examination in convergent light a condensing lens has to be inserted in the opening in the stage directly under the crystal and an auxiliary eyepiece has to be used. This enables us to focus not on the crystal but on the interference figure, which is formed in another plane.

EXAMPLES.

(1) It is found when sodium light is incident on the surface of a certain glass plate at an angle of 58° 18′, that the reflected light is plane polarised. What is the refractive index of the glass?

(2) A 30° Iceland spar prism is cut with its axis parallel to the refracting edge, and light from a sodium flame is incident normally on one of the faces. Calculate the deviation for the ordinary and extraordinary rays.

(3) A 60° quartz prism is cut with its axis parallel to the refracting edge and is placed on the table of a spectrometer. The slit is illuminated by a sodium flame. What are the angles of minimum deviation for the ordinary and extraordinary rays?

(4) A ray of light is incident on the surface of an Iceland spar crystal at an angle of 60° with the normal. The axis is in the surface and in the plane of incidence. What are the angles of refraction of the ordinary and extraordinary rays?

(5) A quartz wedge, the angle of which is ½°, is cut with its optic axis parallel to the edge. It is viewed with a Norremberg's apparatus in such a way that the plane polarised light from a sodium flame passes perpendicularly through the wedge, is reflected by a mirror, and then retraces its path through the wedge to the analysing nicol. What is the distance between the interference bands?

(6) A plate of Iceland spar 2 mm. thick is cut perpendicularly to the axis. Calculate in angular measure the diameters of the first three rings seen when a converging beam of plane polarised light is transmitted through it in the direction of the axis and then analysed by a nicol.

(7) A plate of Iceland spar of thickness d is cut at right angles to the optic axis. A mark is scratched on one face and the plate is then placed on the stage of a microscope with this face down, and the microscope focussed on the scratch. It is found that the latter can be seen sharply in two positions, which are apparently d_o and d_e cms. respectively below the upper surface of the plate. Show that $\mu_o = d/d_o$ and $\mu_e = d/\sqrt{d_o d_e}$, μ_o and μ_e being the two principal indices of refraction of Iceland spar. (Note that the radius of curvature of the extraordinary wave-front is $\mu_o^2 d/\mu_e^2$ where it meets the upper surface of the plate.)

CHAPTER XII.

THE PROPAGATION OF LIGHT IN CRYSTALS.

In the previous chapter only uniaxal crystals have been treated. The manner of the propagation of light in optically biaxal crystals can best be described with reference to a surface,

$$a^2x^2 + b^2y^2 + c^2z^2 = V^2,$$

called the ellipsoid of elasticity. In the equation the coordinate axes are taken in fixed directions in the crystal, a, b, and c are constants depending on the nature of the crystal, and V is the velocity of light in vacuo. The ellipsoid of elasticity was introduced by Fresnel on theoretical considerations, but here it will be considered merely as the simplest way of describing the facts. It has the following property:—

Take any plane through the centre of the ellipsoid. It intersects the ellipsoid in an ellipse. Two waves are propagated in the crystal in the direction at right angles to this plane. Their velocities are numerically equal to V/r_1 and V/r_2 where r_1 and r_2 are the lengths of the major and minor semi-axes of the ellipse. Also each wave is plane polarised in the plane at right angles to the semi-axis that specifies its velocity.

The two waves are thus plane polarised at right angles to one another. Also the construction gives wave velocities, not ray velocities. As the inclination of the plane of section alters, the lengths of both axes of the ellipse vary. Thus neither of the two waves has a constant velocity and neither obeys the ordinary laws of refraction. The distinction between ordinary and extraordinary wave breaks down; both waves are extraordinary. If Ox is the direction of wave propagation, the equation to the ellipse of section is

$$b^2y^2 + c^2z^2 = V^2,$$

r_1 and r_2 equal V/b and V/c, and the velocities of the two waves are consequently b and c. Similarly c and a are the velocities of the two waves propagated in the Oy direction and a and b the velocities of the two waves propagated in the Oz direction. The ratios V/a, V/b, and V/c are called the principal refractive indices of the crystal. When they and the directions for which they are determined are known, the optical properties of the crystal are fully known. We shall suppose in what follows that $a > b > c$.

Optic Axis. The equation

$$x^2 + y^2 + z^2 = r^2$$

represents a sphere of radius r. It is shown in text-books on solid geometry that the equation

$$a^2x^2 + b^2y^2 + c^2z^2 = \frac{V^2}{r^2}(x^2 + y^2 + z^2)$$

or

$$x^2\left(a^2 - \frac{V^2}{r^2}\right) + y^2\left(b^2 - \frac{V^2}{r^2}\right) + z^2\left(c^2 - \frac{V^2}{r^2}\right) = 0$$

represents a cone through the origin and the curve of intersection of the sphere and ellipsoid. If $r = V/b$ the cone reduces to two planes,

$$x^2(a^2 - b^2) = z^2(b^2 - c^2),$$

and these planes cut the ellipsoid in circles.

It follows that there are two directions for which both waves have the same velocity, that velocity being equal to b. These directions are in the xz plane and equally inclined to Oz. They are called the optic axes. In general there are two and hence the crystal is said to be optically biaxal, but if we write $a = b$ they both coincide with Oz and the crystal becomes optically uniaxal. If the wave normal is in an optic axis, the construction for the plane of polarisation becomes indeterminate and consequently the latter may have any direction.

Dispersion of the Optic Axes. In most biaxal crystals the directions of the optic axes vary with the wave-length. Thus in an extreme case, brookite, the plane of the optic axes rotates through 90° in going from one end of the visible spectrum to the other. When selenite is heated it becomes at first uniaxal, then on raising the temperature still higher it becomes biaxal again but with the plane of the optic axes at right angles to its original direction. The change is reversed on cooling, except when the heating has been very prolonged when the reverse change stops at the uniaxal stage.

Fresnel's Law for the Velocity. We shall now derive an equation giving the velocity in any direction as a function of (l, m, n) the direction cosines of that direction.

The plane

$$lx + my + nz = 0$$

cuts the cone

$$x^2\left(a^2 - \frac{V^2}{r^2}\right) + y^2\left(b^2 - \frac{V^2}{r^2}\right) + z^2\left(c^2 - \frac{V^2}{r^2}\right) = 0$$

in two straight lines through the origin, and the distance from the origin to the surface of the ellipsoid of elasticity measured along these lines is r. If the two lines coincide, the plane touches the cone and r becomes the length of a semi-axis of the ellipse in which the plane intersects the ellipsoid of elasticity. The condition of tangency is

$$\frac{l^2}{a^2 - \frac{V^2}{r^2}} + \frac{m^2}{b^2 - \frac{V^2}{r^2}} + \frac{n^2}{c^2 - \frac{V^2}{r^2}} = 0.$$

But if we write v for the velocity in the direction (l, m, n), $v = V/r$. Hence

$$\frac{l^2}{a^2 - v^2} + \frac{m^2}{b^2 - v^2} + \frac{n^2}{c^2 - v^2} = 0 \quad . \quad . \quad . \quad (21)$$

This is the required equation. It is due to Fresnel. As was to be expected, it is a quadratic for v^2.

The Wave-Surface. Let us suppose that a light disturbance travels out in all directions from a point inside a biaxal crystal and that it is required to find the shape of the wave-surface. Take the point as origin and let a great number of plane wave-fronts travel through it in all possible directions. Let FG, HK, and LM be the traces after unit time of wave-fronts, the directions of propagation of which make small angles with one another. These wave-fronts all touch the wave-surface, so the latter is the envelope of all the plane wave-fronts.

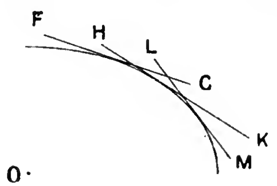


FIG. 192.

If we are given a family of surfaces represented by

$$f(x, y, z, a, b) = 0,$$

the ordinary method of finding its envelope is to form the two equations

$$\frac{\partial}{\partial a} f(x, y, z, a, b) = 0, \quad \frac{\partial}{\partial b} f(x, y, z, a, b) = 0, \quad . \quad (22)$$

and by their means eliminate a and b from the original equation for the family of surfaces. Here the family of surfaces is

$$lx + my + nz = v, \quad . \quad . \quad . \quad . \quad (23)$$

where v is a function of l, m, n given by (21). Only two of the three variables l, m, n are independent, since, of course, they are connected by the equation

$$l^2 + m^2 + n^2 = 1, \quad . \quad . \quad . \quad . \quad (24)$$

and it is shorter to proceed otherwise, by the method of undetermined multipliers.

From (23), (24), and (21) we have

$$xdl + ydm + zdn = dv, \quad . \quad . \quad . \quad (25)$$

$$ldl + mdm + ndn = 0, \quad . \quad . \quad . \quad . \quad (26)$$

and

$$\frac{ldl}{v^2 - a^2} + \frac{mdm}{v^2 - b^2} + \frac{ndn}{v^2 - c^2} = Kvdv, \quad . \quad . \quad (27)$$

where

$$K = \frac{l^2}{(v^2 - a^2)^2} + \frac{m^2}{(v^2 - b^2)^2} + \frac{n^2}{(v^2 - c^2)^2}.$$

The three equations (25), (26), and (27) are equivalent to the two independent equations (22). Hence it is possible to find quantities, such that when (26) and (27) are multiplied by them and added, the coefficients in the resultant equation are equal to the coefficients in (25). Let A and B be such quantities; then

$$\left.\begin{aligned} x &= \mathsf{A}l + \mathsf{B}\frac{l}{v^2 - a^2} \\ y &= \mathsf{A}m + \mathsf{B}\frac{m}{v^2 - b^2} \\ z &= \mathsf{A}n + \mathsf{B}\frac{n}{v^2 - c^2} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (28)$$

$$1 = \mathsf{BK}v \quad . \quad . \quad . \quad . \quad . \quad (29)$$

We have now to eliminate A and B. Multiply (28) by l, m, n respectively, add, and we obtain by means of (23), (24) and (21),

$$v = \mathsf{A} \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

Squaring and adding the two sides of the different components of (28) we have

$$r^2 = \mathsf{A}^2 + \mathsf{B}^2\mathsf{K},$$

the product term disappearing on account of (21). On combining this with (29) and (30) we obtain

$$\mathsf{B} = \mathsf{B}^2\mathsf{K}v = (r^2 - v^2)v.$$

Substitution for A and B in (28) gives

$$x = vl + \frac{v(r^2 - v^2)l}{v^2 - a^2} = \frac{vl(r^2 - a^2)}{v^2 - a^2},$$

or $\qquad \dfrac{x}{r^2 - a^2} = \dfrac{vl}{v^2 - a^2} = \dfrac{x - vl}{r^2 - v^2}$

and similarly $\qquad \dfrac{y}{r^2 - b^2} = \dfrac{vm}{v^2 - b^2} = \dfrac{y - vm}{r^2 - v^2} \qquad . \quad . \quad . \quad (31)$

and $\qquad \dfrac{z}{r^2 - c^2} = \dfrac{vn}{v^2 - c^2} = \dfrac{z - vn}{r^2 - v^2}$

On multiplying these equations by x, y, z respectively and adding we obtain the required equation of the wave-surface, namely

$$\frac{x^2}{r^2 - a^2} + \frac{y^2}{r^2 - b^2} + \frac{z^2}{r^2 - c^2} = 1 \quad . \quad . \quad . \quad (32)$$

It gives the position of the wave-surface after unit time.

Sections of the Wave-Surface by the Co-ordinate Planes. The simplest way of getting an idea of the shape of the wave-surface is to consider its sections by the three coordinate planes. If we get rid of the denominator (32) becomes

$$x^2(r^2 - b^2)(r^2 - c^2) + y^2(r^2 - c^2)(r^2 - a^2) + z^2(r^2 - a^2)(r^2 - b^2)$$
$$= (r^2 - a^2)(r^2 - b^2)(r^2 - c^2) \quad . \quad . \quad . \quad (33)$$

When $x = 0$, this gives

$$(r^2 - a^2)\left(y^2(r^2 - c^2) + z^2(r^2 - b^2) - (r^2 - b^2)(r^2 - c^2)\right) = 0.$$

The second factor simplifies to

$$- y^2c^2 - z^2b^2 + (y^2 + z^2)(b^2 + c^2) - b^2c^2 = 0,$$

or $\qquad y^2b^2 + z^2c^2 - b^2c^2 = 0.$

Hence the yz plane intersects the wave-surface in

$$y^2 + z^2 = a^2$$

and
$$\frac{y^2}{c^2} + \frac{z^2}{b^2} = 1,$$

a circle and an ellipse lying wholly inside it.

When $y = 0$, (33) reduces to

$$(r^2 - b^2)\,(x^2(r^2 - c^2) + z^2(r^2 - a^2) - (r^2 - a^2)\,(r^2 - c^2)\,) = 0.$$

The second factor simplifies to $x^2a^2 + z^2c^2 - a^2c^2 = 0$.
Hence the zx plane intersects the wave-surface in

$$x^2 + z^2 = b^2$$

and
$$\frac{x^2}{c^2} + \frac{z^2}{a^2} = 1,$$

a circle and an ellipse which intersect one another.

When $z = 0$, (33) reduces to

$$(r^2 - c^2)\,(x^2(r^2 - b^2) + y^2(r^2 - a^2) - (r^2 - a^2)\,(r^2 - b^2)\,) = 0.$$

The second factor simplifies to $x^2a^2 + y^2b^2 - a^2b^2 = 0$.
Hence the xy plane intersects the wave-surface in

$$x^2 + y^2 = c^2$$

and
$$\frac{x^2}{b^2} + \frac{y^2}{a^2} = 1,$$

an ellipse with a circle inside it.

The sections are represented by the following diagrams (fig. 193).

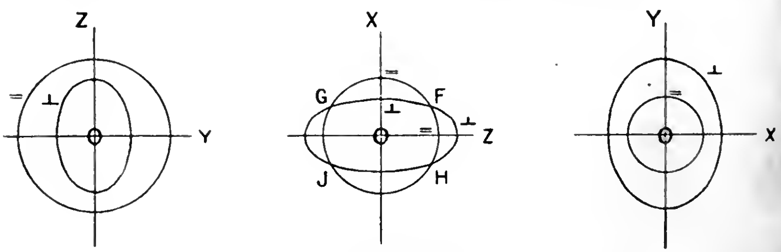


FIG. 193.

The direction of polarisation can be obtained from the ellipsoid of elasticity and is shown in the diagram, ⊥ denoting that the light is polarised at right angles to the plane of the figure, and = denoting that it is polarised in the plane of the figure.

The wave-surface, then, consists of two sheets which meet in four points F, G, J, and H and nowhere else. These points lie two and two on straight lines through the origin called the axes of single ray velocity. The axes of single ray velocity are quite distinct from the optic axes.

When a wave is refracted at the surface of a biaxal crystal, the refracted ray and wave-front can be obtained from the wave-surface by Huygens' construction just as in the case of a uniaxal crystal. In the case of a biaxal crystal, however, conditions are more complex; in general neither of the refracted rays is in the plane of incidence.

Although the two waves belonging to the same normal are plane polarised at right angles to one another, the two planes of polarisation belonging to a given ray are not perpendicular to one another unless the ray coincides with the wave normal.

Internal Conical Refraction. Let fig. 194 represent the section of the wave-front by the zx plane. Then it is obvious we can have a plane touching one sheet of the surface at M and the other at N. This plane is perpendicular to the radius joining M to the origin, which is of course equal to b. Since the tangent planes to the two sheets perpendicular to this radius coincide, it is the optic axis.

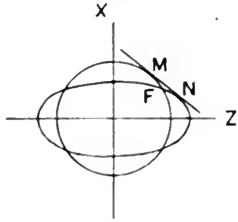


FIG. 194.

It can be shown that the common tangent plane meets the surface not only in the two points M and N but also in a circle of which MN is a diameter. For from (31), multiplying the first equation by l and the third by n, we have

$$\frac{lx}{r^2 - a^2} + \frac{nz}{r^2 - c^2} = v\left(\frac{l^2}{v^2 - a^2} + \frac{n^2}{v^2 - c^2}\right).$$

If (l, m, n) refer to the optic axis, $l^2/(a^2 - b^2) = n^2/(b^2 - c^2)$, $m = 0$, and $v = b$.

Hence the right-hand side of the above equation

$$= v\left(\frac{l^2}{b^2 - a^2} + \frac{n^2}{b^2 - c^2}\right) = 0,$$

and $$lx(r^2 - c^2) + nz(r^2 - a^2) = 0 \quad . \quad . \quad . \quad . \quad (34)$$

In this equation x, y, z determine the point of contact of the ray with the wave-front in the direction (l, m, n). The equation to the tangent plane at M is

$$lx + nz = b \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

On combining (34) and (35) we obtain

$$b(x^2 + y^2 + z^2) - lc^2x - na^2z = 0, \quad . \quad . \quad (36)$$

the equation to a sphere passing through the origin.

The locus of the point of contact is therefore given by the intersection of (35) and (36), a sphere and plane, and must be a circle. At the point F there is a little pit or depression in the surface; the tangent plane MN covers it over and touches the surface all round it in a circle. As the direction of the ray is given by the point of contact of the tangent plane, it follows that in this case there is an infinite number of rays joining the origin to the circle and lying on the surface of a cone.

This result was derived theoretically by Sir William Hamilton, and was verified experimentally at his request by Dr. Lloyd. A plate of arragonite with its faces cut perpendicular to the bisector of the angle between the optic axes was used, that is, a plate with its faces parallel to the xy coordinate plane. For this crystal the vertical angle of the cone is comparatively large. A thin pencil of light SO was incident on the plate through a small hole in a screen at O. The emergent rays were received on a screen EF at the other side. The angle of incidence of SO was varied very slowly until the refracted wave normal coincided with the optic axis; then the two images on the screen EF extended into a luminous ring of light. The rays traversed the plate as a cone and on emerging formed a hollow cylinder, the sides of which were parallel to SO.

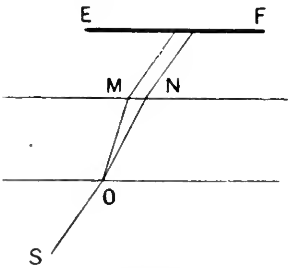


FIG. 195.

External Conical Refraction. An infinite number of tangent planes can be drawn to the wave-surface at the vertex of the conical depression where the two sheets meet. These tangent planes envelop a cone. If a ray traverses a crystal plate in the axis of single ray velocity, to that ray there corresponds an infinite number of wave normals, all lying on a cone. When the ray reaches the surface of the plate, the direction of the emergent ray depends on the wave normal; consequently the single ray gives rise to an infinite number of emergent rays, which all lie on the surface of a cone.

This phenomenon, which is known as external conical refraction, was also predicted by Sir William Hamilton and verified experimentally by Dr. Lloyd. A plate of arragonite was used, and on each face a plate of metal with a very fine hole was placed. The plates were adjusted so that the line OP (fig. 196) joining the holes coincided with the axis of single ray velocity. Then, when light was incident on the one hole in a suitable direction and the observer looked into the other, a bright ring of light was seen. As the emergent rays are parallel to the incident rays, to obtain the full emergent cone the incident pencil must also be conical and have at least as great a vertical angle as the emergent cone. Of course it is not necessary for it to be a hollow cone; the other rays do not travel along the axis of single ray velocity and are consequently stopped by the plate on the other face.

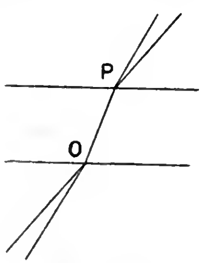


FIG. 196.

Interference Phenomena in Biaxal Crystals. Figs. 197 and 198 represent interference phenomena produced in convergent monochromatic light by a biaxal crystal plate which has its faces cut perpendicular to the bisector of the angle between the optic axes. Both figs. are for the case of crossed nicols; in fig. 197 the plane of the axes is parallel to the principal plane of one of the nicols, and in fig. 198 it makes an angle of 45° with the principal planes of the nicols.

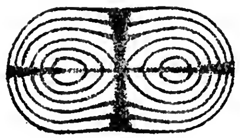


FIG. 197.

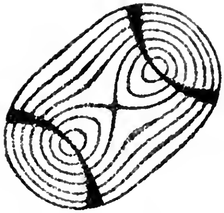


FIG. 198.

Many crystalline bodies, including rock salt (NaCl) and sylvin (KCl), are optically isotropic and behave in exactly the same way as glass as regards the propagation of light through them.

EXAMPLES.

(1) Mica is a biaxal crystal which cleaves naturally in planes perpendicular to the bisector of the angle between the optic axes. Its principal indices of refraction for sodium light are 1·5601, 1·5936, and 1·5977. Calculate the angle which each optic axis makes with the normal to the plane of cleavage. (The principal indices of refraction of mica vary considerably from specimen to specimen; some samples are almost uniaxal.)

(2) Calculate the velocities of sodium light in mica in a direction perpendicular to the cleavage planes. Hence calculate the thickness of a quarter-wave plate and of a half-wave plate made from mica. (A quarter-wave plate is one, in passing through which the faster ray gains one-quarter period on the slower ray; a half-wave plate is one, in passing through which the faster ray gains one-half period on the slower ray.)

(3) Selenite ($CaSO_4 + 2H_2O$) is a biaxal crystal which cleaves parallel to the plane containing its optic axes. How thick should a quarter-wave plate of selenite be, if its faces are cleavage planes? It is to be constructed for sodium light, and the principal refractive indices of selenite for sodium light are 1·530, 1·523, and 1·521.

(4) A mica plate has its surfaces parallel to the cleavage planes. It is introduced into the path of a beam of convergent light between crossed nicols and gives interference figures similar to those shown in fig. 197. The plate is adjusted until one of the "eyes," i.e. the direction of one of the axes, appears in the centre of the field. It is then rotated through an angle 2θ and the other "eye" brought into the centre of the field. This angle 2θ is called the apparent angle between the optic axes. How is it related to the true angle which has to be calculated in ex. 1? (Hint: correct for refraction.)

CHAPTER XIII.

OPTICAL ROTATION AND THE ANALYSIS OF POLARISED LIGHT.

Suppose that two nicols, a polariser and an analyser, are mounted so as to rotate about the same axis, and that a tube filled with water and with plane glass ends is placed between them, so that the rays from the polariser pass down it before entering the analyser. Then when the two nicols are crossed, the field is dark. If, however, some sugar is dissolved in the water, as it goes into solution, the field becomes light, but extinction can always be restored by rotating the analyser through a definite angle. The sugar solution thus rotates the plane of polarisation of the incident light. If the strength of the solution is constant, the angle of rotation is proportional to the length of the tube. For a given length of tube the rotation is proportional to the strength of the solution measured in grams of dissolved substance per c.c. of solution. The rotation varies with the wave-length, being roughly proportional to the inverse square of the latter. It also varies with the temperature.

Substances which rotate the plane of polarisation as a sugar solution does are said to be optically active. They are divided into two classes according to the direction of the rotation. If to an observer looking towards the source the plane of polarisation is rotated in the direction in which the hands of a clock revolve, the rotation is said to be positive or right-handed or dextrogyric, if in the other direction the rotation is said to be negative or left-handed or laevogyric.

The optical activity of a substance is defined by its specific rotation which is usually written $[\alpha]_t$, the temperature in degrees centigrade being substituted for the suffix t. The specific rotation is the rotation per decimetre of solution divided by the grams of active substance per cubic centimetre of solution. Thus, if 3 gms. of sugar are dissolved in water and the volume of the solution is 80 c.c., if the solution is poured into a tube 22 cm. long and the plane of polarisation of a beam of a sodium light that passes down the tube is rotated through 5·5°, the specific rotation is given by

$$\alpha = \frac{5{\cdot}5 \times 80}{2{\cdot}2 \times 3} = 66{\cdot}7^\circ.$$

The molecular rotation is the specific rotation multiplied by the molecular weight of the dissolved substance. a varies with the wave-length, with the nature of the solvent, and with the concentration of the solution.

The optical activity of liquids was observed first by Biot for sugar solutions in 1815. It was found by Arago in 1811 that quartz rotates the plane of polarisation of a beam traversing it in the direction of the optical axis, and the same fact was afterwards observed for many other crystals. In the case of quartz the rotation is simply given in degrees per mm. of thickness; the value for sodium light and 20° C. is 21·72° per mm. For some specimens of quartz the rotation is right-handed and for others it is left-handed, but the numerical value is always the same.

Fused quartz is inactive, i.e. does not rotate the plane of polarisation at all, so the rotation is due to the crystalline structure, to the arrangement of the molecules with reference to one another. We are accustomed to think of liquids as perfectly isotropic and to assume that in their case the molecules move freely past one another. Hence in liquids the cause of the optical activity must lie in the structure of the molecules themselves. This view has received a great development on the chemical side. The liquids and dissolved solids, which are optically active, possess in almost every case an asymmetric atom of carbon, tin, sulphur, or nitrogen, and to each of them there corresponds a twin substance, which rotates the plane of polarisation through the same angle in the opposite direction. The carbon atom is said to be asymmetric if the four radicals attached to it are all different. If we

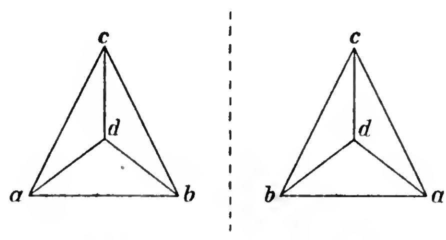


FIG. 199. FIG. 200.

represent the molecule in space by attaching the radicals a, b, c, and d at the corners of a tetrahedron, then we have the two possible arrangements shown in figs. 199 and 200. In these figures the corner d is supposed to be in each case above the plane of the paper. It will be found by trial that any other arrangement can be made to coincide with one of the above merely by turning the tetrahedron round. These two arrangements cannot, however, be made to coincide with each other, for if c and d in the first figure coincide with c and d in the second, a of the first falls on b of the second. The one arrangement is the mirror image of the other, the position of the mirror being shown by the dotted line. The rotation is supposed to be due to the fact that the molecule has this asymmetric shape, the one figure giving the dextro compound and the other the laevo compound, but, though this idea has been of great help in elucidating chemical structure, it has not yet been possible to explain why or how the asymmetric shape affects the light wave in this way. The two compounds represented in figs. 199 and 200 are said to be optical isomers.

Saccharimetry. The rotation of the plane of polarisation is used in commerce and also in medicine as a means of estimating the quantity of sugar in a liquid. The molecular rotation of a substance is also very important theoretically. Consequently a number of instruments have been designed for measuring optical rotation. Such instruments are called saccharimeters or polarimeters.

The simplest form of saccharimeter, that used by Mitscherlich, consists only of two nicols, a polariser and an analyser, which rotate about the same axis, and a tube between for holding the liquid. The analyser is set at extinction, first when the tube is empty and then when it is filled with the liquid. The difference between the two positions gives the required angle but only to a multiple of π. For example, if the analyser had to make a right-handed rotation of θ in order to restore extinction, then the actual rotation may be $\pi + \theta$ or, if it is a left-handed one, $\pi - \theta$. The correct sign and multiple of π, if any, can be found from the fact, that if the tube is half the length or the solution half as strong the value of the rotation diminishes by half. Also, if instead of sodium light red light is used, the rotation usually diminishes in the ratio of about five to four. A suitable red light for a rough test of this kind may be obtained by putting a piece of red glass in front of a white flame.

A simple saccharimeter of this type is, however, very inaccurate on account of the difficulty of setting the analyser at extinction. The field appears perfectly dark not for a single position but for a region. The usual way of working is to rotate the analyser slowly through this region until the field begins to be bright again, then to turn it back to equal brightness on the first side, and finally to rotate it half way into the region, the angle being estimated with the fingers. The reading is then taken. But this method is inaccurate at its best.

Modern instruments are constructed on what is known as the "half shadow" principle. Fig. 201 represents the optical system of one of the most usual types, one fitted with a Lippich two-prism polariser.

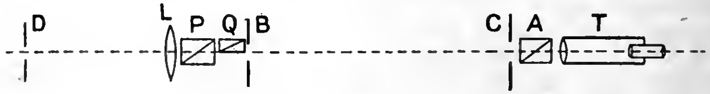


FIG. 201.

The Lippich polariser consists of the two nicols, P and Q, and an aperture B. A is an analyser which is placed behind an aperture C. The analyser of course rotates about the axis of the instrument and its position can be read on a divided circle. T is a small astronomical telescope which is focussed on the edge of Q. The tube to be examined is placed between B and C. D is an aperture for admitting the light and L is a convex lens. The sodium flame used as source is placed at such a distance from D, that, when the tube is in, its image falls on the aperture C.

On looking into the telescope we see the field divided into two halves of unequal brightness separated by a very sharp line (fig. 202). The sharp line is the image of the edge of the prism Q; the one half of the field is illuminated by light which passes through both the prisms P and Q, the other half by light which passes through the prism P alone. The principal planes of P and Q are inclined at a small angle δ to one another; they are represented by the arrows p and q. If the principal plane of the analyser is at right angles to p, one half of the field is black, if at right angles to q the other half is extinguished, and, if we turn the analyser from one of these positions to the other, the intensity of the one half of the field rapidly increases from zero and of the other half of the field rapidly decreases to zero. Consequently for an intermediate position the two halves are equally bright. This is the position that we set on. For it the principal plane of the analyser has the direction given by the dotted line.

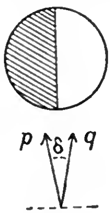


FIG. 202.

The smaller the angle δ, the greater is the change in the relative intensity of the two halves of the field produced by a definite small rotation of the analyser. If, however, δ is very small, the illumination of the field becomes too faint. There is consequently a definite value of δ for which the accuracy of the setting is a maximum. This value depends on the brightness of the light source and the transparency of the solution, and consequently it is an advantage to be able to adjust δ, so as to get maximum sensitiveness for any given conditions.

In addition to the two-prism polariser Lippich invented a three-prism polariser, in which there are two little nicols set in front of a larger one with their principal planes parallel, and which divides the field into three parts (fig. 203), the centre being illuminated by light that passes through the large nicol and each side by light that passes through the large nicol and one of the small ones. The two sides have the same brightness. This arrangement is used for the most accurate work. With the ordinary half-shadow arrangement, if the eye gets off the axis of the instrument, the halves may match at the wrong angle. This error is impossible with the three-prism polariser, for, if the eye moves to one side and that side appears too bright, the other side appears too weak.

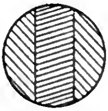


FIG. 203.

Besides the Lippich two-prism polariser there are other older methods of producing a half-shadow field, for example, the Laurent plate and the biquartz. The Laurent plate can be used only for one special wave-length. It consists of a quartz plate cut with the axis in the surface parallel to one edge and of such a thickness that the ordinary wave gains $\lambda/2$ on the extraordinary wave in passing through it. It is placed immediately after the polarising nicol and covers half the field. The other half of the field is covered by a glass plate that transmits the same amount of light as the quartz. Let the axis of the quartz make an angle ϕ with the principal plane of the polarising nicol. Then if OB gives the direction of the axis of the plate and if OP represents the incident vibration, on entering the plate the latter decomposes into OA and OB. These two components suffer a relative phase difference $\lambda/2$ in passing through the plate; hence after emerging, if the one component is represented by OB the other must be represented by OA′ and the resultant by OP′. The insertion of the plate has therefore rotated the plane of polarisation of the light that passes through it by 2ϕ.

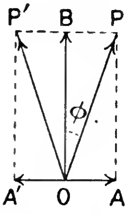


FIG. 204.

A plate cut from a uniaxal crystal with its faces parallel to the axis and producing a phase difference of half a wavelength between the two rays, as the quartz plate does in Laurent's arrangement, is called a half-wave plate.

The biquartz is placed immediately after the polariser. It consists of two semicircular pieces of quartz cut with their surfaces perpendicular to the axis, about 3·75 mm. thick, the one piece being right-handed and the other left-handed. They are cemented together along the diameter. If sodium light is used, each half of the field is rotated through about 80°, so that they are inclined at 20° to one another. The analyser is adjusted to make both halves of the field equally bright. If the rotation to be measured is small, the biquartz can also be used with white light, for then for each constituent colour both halves are matched at the one position of the analyser. In this case, when the analyser is turned through the correct position, the colour of one half of the field changes suddenly from blue through a greyish-violet called the sensitive tint to pink, and of the other half from pink through greyish-violet to blue, and so the setting can be made with considerable accuracy. When the setting is made, the colour for which the rotation is

$$\frac{90^\circ}{3\cdot 75} = 24^\circ,$$

i.e. for which λ is about $\cdot 56\ 10^{-4}$ cm., is extinguished, the middle of the spectrum is cut out and the illumination of the field is due to the end colours red and blue, which together give grey. If the analyser is rotated slightly from this position, the missing part in the spectrum moves towards the red for the one half of the field and towards the blue for the other, and consequently in each case the other colour preponderates in the mixture.

In some saccharimeters the analyser does not rotate and the rotation produced by the solution is measured by screwing in a quartz wedge to remove it. The quartz produces an equal and opposite rotation to that produced by the solution. This method is specially suitable for sugar solutions, for the dispersion of the rotation, i.e. the variation of the rotation with wave-length is nearly the same for quartz and sugar, and consequently white light can be used. Soleil's saccharimeter, which employs this method, is represented in fig. 205. The light on

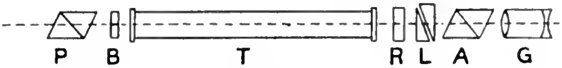


Fig. 205.

entering the instrument first passes through the polarising nicol P, then the biquartz B and tube T for holding the solution. It next passes through a plate of right-handed quartz, R, cut with its surfaces perpendicular to the axis, and through two wedges of left-handed quartz, L. These wedges have the same angle, so together they form a plate of left-handed quartz of variable thickness. They are cut with their axes perpendicular to the surfaces of this plate. Their angles are, of course, very much exaggerated in the diagram. A is the analyser, which is fixed so as to produce the tint of passage when the tube is empty and the thickness of the wedges exactly equal to the thickness of the plate R. G is a small Galilean telescope which is focussed on the biquartz. If the solution produces a right-handed rotation, the tint of passage is restored by screwing in the second wedge, and if it produces a left-handed rotation, by screwing out the same wedge. The position of the wedge is read on a scale which is calibrated in terms of degrees.

For accurate saccharimetry it is necessary that the light should be bright and monochromatic. Usually observations are made only for sodium light and special burners are sold for producing the yellow flame, but none of these are better than the simple sodium bicarbonate bead in a loop of fine platinum wire. When the sodium yellow is extinguished by the analyser, the blue background of the bunsen flame becomes troublesome; to absorb it a glass cell containing an aqueous solution of potassium bichromate is used in front of the aperture of the instrument. Or instead of using a filter a direct vision spectroscope may be screwed on in front of the eyepiece and a slit fixed immediately in front of the Lippich polariser. This arrangement draws the field out into a spectrum parallel to the dividing line between the two halves of the field and so separates the sodium yellow from the blue background. Also if white light or, better still, if the mercury arc is used as source, this arrangement enables determinations of the optical activity to be made in different parts of the spectrum. The mercury arc permits of the use of a wider slit and a more accurate determination of the wave-length.

Fig. 206 represents a polarimeter to which a Lippich polariser or

Laurent plate or biquartz or Soleil wedges can be fitted. In the

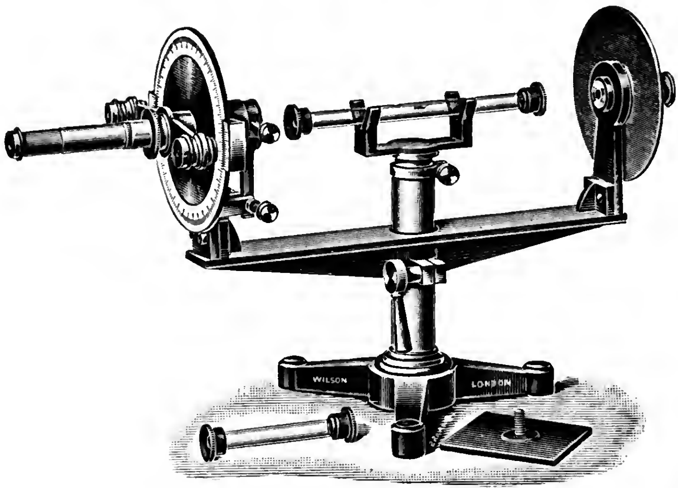


FIG. 206.

figure it is shown only with polariser and analyser and with a tube in position.

Determination of the Sugar in a Solution when other Active Substances are present. An aqueous solution of right-handed cane sugar, $C_{12}H_{22}O_{11}$, is changed into left-handed invert sugar, which is a mixture of equal parts of dextrose and levulose, on heating with hydrochloric acid for 10 minutes at 70° C. Dextrose and levulose have the same formula, $C_6H_{12}O_6$. This reaction enables us to determine the amount of cane sugar present in a solution when it is mixed with other active substances. After the rotation with sodium light, θ, has been determined for the solution, 50 c.cs. of the latter are measured off, 5 c.cs. of strong hydrochloric acid added, and the mixture kept at a temperature of 70° C. for about 10 minutes. The concentration in gms./c.c. is of course now only 10/11 of the original value. The solution is then cooled, its temperature noted, and the rotation determined and multiplied by 1·1. Let the result be θ'.

Let c be the concentration of the cane sugar in gms. per 100 c.cs. of solution, and let ϕ be the part of the rotation produced by substances other than cane sugar. The specific rotation of cane sugar is 66·5; hence

$$\theta - \phi = 66{\cdot}5 \frac{cl}{100} \quad . \quad . \quad . \quad . \quad (37)$$

where l is the length of the tube in decimetres. The specific rotation of invert sugar is given by

$$-19{\cdot}7 - {\cdot}036\, c' + {\cdot}304\,(t - 20),$$

where c' is the concentration of the invert sugar in gms. per 100 c.c. of solution and t is the temperature in degrees centigrade. The molecular weight of cane sugar is 342; after conversion to invert sugar it becomes 360;

hence $$c' = \frac{360}{342} c.$$

We have $$\theta' - \phi = \{ - 19{\cdot}7 - {\cdot}036\, c' + {\cdot}304\,(t - 20)\} \frac{lc'}{100} \quad . \quad (38)$$

The subtraction of (38) from (37) gives

$$\theta - \theta' = 66{\cdot}5 \frac{cl}{100} - \{ - 19{\cdot}7 - {\cdot}036 \times 1{\cdot}053\, c + {\cdot}304\,(t - 20)\} \frac{lc\, 1{\cdot}053}{100},$$

whence c can be determined. The term involving c inside the bracket is small; hence we can ignore it at first and so find a rough value for c, and then substitute this rough value in it and thus find a more accurate value. The specific rotation of cane sugar does not vary nearly so much with temperature as the specific rotation of invert sugar does.

Magnetic Rotation. Faraday discovered in 1845 that isotropic substances of high refractive index rotate the plane of polarisation of plane polarised light when placed in a strong magnetic field. This phenomenon is known as the Faraday effect. In his experimental arrangement the iron pole pieces of a large magnet were bored, the two holes being in the same straight line so that a beam of light could pass down the one, then through the most intense part of the field parallel to the lines of force and finally through the other. The beam of light passed through a polarising nicol before entering the magnet and through an analysing nicol after emerging from it, and a block of dense lead glass was placed between the poles in the path of the beam. When the current passed through the coils of the magnet and a magnetic field was produced the plane of polarisation was rotated. As long as the field strength remained constant, the rotation remained constant. When the field was reversed the rotation was reversed.

The rotation is proportional to H, the strength of the magnetic field in gausses, and to l, the length of path traversed in it in cms., i.e. if θ denotes the rotation in minutes, then

$$\theta = r\mathrm{H}l,$$

where r is a constant. This constant is known as Verdet's constant and it is roughly inversely proportional to the square of the wavelength. It may be either positive or negative. If the rotation is in the direction of the electric current producing the field, that is, if it appears clockwise to an observer looking in the direction of the field, then it is said to be positive. In the other case it is said to be negative. The rotation thus does not depend on the direction of the light, and we have an important difference between magnetic rotation and the natural rotation occurring in quartz and sugar.

In the case of a cane sugar solution, for example, the rotation was right-handed to an observer looking towards the source. Hence if a mirror is placed at the end of the tube and the beam made to retrace its path, to obtain the rotation produced by the second passage the observer must look the other way and consequently the one rotation annuls the other. In the case of magnetic rotation, if the light passes through the substance in the direction of the field and is reflected by a mirror so as to retrace its path, since the direction of the field does not change, if the rotation appears right-handed to an observer looking in the direction of the light during the first passage, it appears left-handed to an observer looking in the direction of the light during the second passage. Thus by reflecting the light the rotation is doubled. Magnetic rotation can be multiplied by reflecting the light back and forward through the field; in the case of natural rotation each successive passage annuls the effect of the previous one. This distinction is to be expected, for in natural rotation we are dealing with a "twisted" structure in the substance produced by the passage of the light wave, whereas in magnetic rotation we are dealing with a twisted structure or a rotation in the molecules produced by the magnetic field and consequently depending only on the latter.

Since Faraday's time magnetic rotation has been detected for a great number of substances, including gases, and for the great majority Verdet's constant is positive. For water it has the value ·013, for carbon bisulphide ·043, and for the heaviest flint glass ·0888.

The magnetic rotation of liquids has been studied also from the chemical side. In this connection it is measured by the specific magnetic rotation, which is the ratio of the rotation to the rotation produced by the same thickness of water in the same field divided by the strength of the solution in gms. of dissolved substance per c.c. of solution. The molecular rotation is this quantity multiplied by the molecular weight of the dissolved substance and divided by the molecular weight of water. Molecular rotation is supposed to be additive but with a strong constitutive influence.

An enormous positive rotation is found to take place when light is transmitted through very thin films of iron, nickel, and cobalt placed at right angles to a magnetic field. In this case the rotation is greater at the red end of the spectrum and reaches a limit when the metal is magnetised to saturation. For iron the rotation is a complete revolution for a thickness of ·02 mm.

Elliptically Polarised Light. Consider the expression

$$\eta = b \sin \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right), \quad \zeta = c \sin\left\{\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) + \delta\right\}.$$

If η and ζ represent respectively the displacements parallel to Oy and Oz, the resultant displacement is transverse to the Ox axis. Each of the expressions separately represents a plane polarised wave propagated in the direction Ox, and together they represent the most general type of transverse wave that can be propagated in that direction.

The second expression can be written

$$\zeta = c\left\{\sin \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) \cos \delta + \cos \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) \sin \delta\right\}$$

which, by substitution of the first expression, gives

$$\frac{\zeta}{c} = \frac{\eta}{b} \cos \delta + \cos \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) \sin \delta.$$

This becomes

$$\cos \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) = \frac{\zeta}{c \sin \delta} - \frac{\eta}{b} \cot \delta\,;$$

also

$$\sin \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) = \frac{\eta}{b},$$

hence squaring and adding the right and left-hand sides

$$1 = \left(\frac{\zeta}{c \sin \delta} - \frac{\eta}{b} \cot \delta\right)^2 + \left(\frac{\eta}{b}\right)^2$$

or

$$\frac{\eta^2}{b^2 \sin^2 \delta} - \frac{2\eta\zeta \cos \delta}{bc \sin^2 \delta} + \frac{\zeta^2}{c^2 \sin^2 \delta} = 1 \quad . \quad . \quad (39)$$

This equation represents an ellipse, for it is a central conic and we know from the conditions of the problem that η and ζ are never infinite. We also get the same result on purely mathematical grounds from the fact that the left-hand side equated to zero gives lines parallel to the asymptotes and consequently the asymptotes are imaginary.

Thus during the passage of the wave the vibrating particle describes an ellipse about its position of rest. For this reason the most general type of transverse wave is said to be elliptically polarised.

If the y and z axes are chosen parallel to the major and minor axes of the ellipse, the product terms disappear. Since b and c are always finite, this can happen only by $\cos \delta$ becoming zero, i.e. by δ becoming $\pm \frac{1}{2}\pi$. Thus an elliptically polarised wave referred to its principal axes can always be written in the form

$$\eta = b \sin \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right), \quad \zeta = c \sin\left\{\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) \pm \frac{\pi}{2}\right\} = \pm\, c \cos \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right).$$

If the upper sign is used, to an observer facing the approaching wave the particle appears to run round the ellipse in the same direction as the hands of a clock and the ellipse is said to be right-handed. If the lower sign is used, the particle appears to run round the ellipse in the contrary direction and the ellipse is said to be left-handed.

If $b = c$ the ellipse degenerates into a circle and the wave is said to be circularly polarised, right-handed if the upper sign is used, and left-handed if the lower sign is used.

If either b or c is zero, the wave becomes plane polarised.

Fresnel's Interpretation of the Rotation of the Plane of Polarisation. Consider the equations

$$\eta_1 = a \sin \frac{2\pi}{\tau}\left(t - \frac{x}{v_1}\right),\ \zeta_1 = a \cos \frac{2\pi}{\tau}\left(t - \frac{x}{v_1}\right)$$

and
$$\eta_2 = a \sin \frac{2\pi}{\tau}\left(t - \frac{x}{v_2}\right),\ \ \zeta_2 = -a \cos \frac{2\pi}{\tau}\left(t - \frac{x}{v_2}\right).$$

The first two represent a right-handed circularly polarised wave travelling with the velocity v_1, and the second two represent a left-handed circularly polarised wave with the same period and same amplitude travelling in the same direction with the velocity v_2. If the two waves are superimposed we obtain

$$\eta = \eta_1 + \eta_2 = a\left[\sin \frac{2\pi}{\tau}\left(t - \frac{x}{v_1}\right) + \sin \frac{2\pi}{\tau}\left(t - \frac{x}{v_2}\right)\right]$$
$$= 2a \sin \frac{2\pi}{\tau}\left(t - \frac{x}{2}\left\{\frac{1}{v_1} + \frac{1}{v_2}\right\}\right) \cos \frac{\pi x}{\tau}\left(\frac{1}{v_2} - \frac{1}{v_1}\right)$$

and
$$\zeta = \zeta_1 + \zeta_2 = a\left[\cos \frac{2\pi}{\tau}\left(t - \frac{x}{v_1}\right) - \cos \frac{2\pi}{\tau}\left(t - \frac{x}{v_2}\right)\right]$$
$$= -2a \sin \frac{2\pi}{\tau}\left(t - \frac{x}{2}\left\{\frac{1}{v_1} + \frac{1}{v_2}\right\}\right) \sin \frac{\pi x}{\tau}\left(\frac{1}{v_2} - \frac{1}{v_1}\right),$$

which give
$$\frac{\eta}{\zeta} = -\cot \frac{\pi x}{\tau}\left(\frac{1}{v_2} - \frac{1}{v_1}\right).$$

As x increases, the cotangent runs through the four quadrants, completing a rotation in the distance given by

$$\frac{2\tau}{\dfrac{1}{v_2} - \dfrac{1}{v_1}}.$$

The two circularly polarised waves thus combine to form a plane polarised wave of amplitude $2a$, the plane of polarisation of which rotates uniformly as the wave proceeds. In one centimetre it rotates through

$$\frac{\pi}{\tau}\left(\frac{1}{v_2} - \frac{1}{v_1}\right) \quad . \quad . \quad . \quad . \quad . \quad (40)$$

radians. If the two circularly polarised waves have the same velocities, $v_1 = v_2$, and the plane of polarisation of the resultant wave remains fixed.

Fresnel was the first to suggest, that in an optically active medium the plane polarised wave was decomposed in the above way into two circularly polarised waves of slightly different velocity. He proved that the explanation was true by arranging and cementing together several quartz prisms in the form of a rectangular parallelipiped. The prisms ABC and BED were right-handed, and the prisms CBD and DEF left-handed, and the optical axis of each prism was perpendicular to the end faces AC and EF. If a plane polarised ray is incident on the face AC and decomposes in the manner indicated above

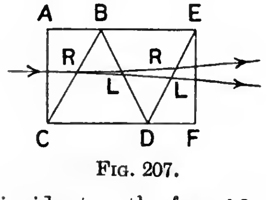


FIG. 207.

into two circular ones, the right-handed wave travels faster, say with velocity v_1 in the first prism, and slower, with velocity v_2 in the second. It has velocity v_1 again in the third prism and velocity v_2 in the fourth; hence it ought to be refracted in the manner indicated in the figure by the ray LL while the left-handed wave ought to be refracted in the manner indicated by the ray RR. Fresnel found that this was the case, that there were two emergent beams, and that both were circularly polarised.

The necessity for Fresnel's experimental arrangement will be seen if we calculate the magnitude of the effect for a Cornu half prism. In modern spectrographs, i.e. instruments used for photographing spectra, the prism is usually made of quartz, because quartz transmits ultra-violet rays that are absorbed by glass, and is of a special type called a Cornu double prism. This consists of two 30° prisms, one of right-handed quartz and the other of left-handed quartz, placed together so as to form a 60° prism and cut with the optic axes perpendicular to the faces in contact. If one of these prisms is placed on the spectrometer table, the slit illuminated with sodium light, and the rays from the collimator allowed to fall perpendicularly on the face that is at right angles to the axis and to pass through the prism to the telescope, theoretically two images of the slit should be seen in the field of the latter. It is easy to calculate what the angular separation of these images should be. A quartz plate 1 mm. thick rotates sodium yellow through 21·7°. On substituting this value in formula (40) we obtain

$$217 = \frac{\pi}{\tau}\left(\frac{1}{v_2} - \frac{1}{v_1}\right)\frac{180}{\pi}.$$

Now $\tau = \lambda/v$ where λ and v are the wave-length and velocity in air. Hence

$$217 = \frac{180}{\lambda}\left(\frac{v}{v_2} - \frac{v}{v_1}\right) = \frac{180}{\lambda}(\mu_2 - \mu_1),$$

or $\mu_2 - \mu_1 = \dfrac{217}{180}\lambda = 7{\cdot}1 \; 10^{-5}$, where μ_1 and μ_2 are the indices of refraction of the two circularly polarised waves. Their mean value may be taken as the ordinary index of refraction of quartz; thus

$$\frac{\mu_1 + \mu_2}{2} = 1{\cdot}5442.$$

We might solve for μ_1 and μ_2, then calculate the deviations of the two waves separately, and their difference would be the required angular separation, but it is shorter to get the separation directly.

Let ϕ be the mean angle of refraction and μ the mean index of refraction. The angle of incidence internally is 30°; hence

$$\sin\phi = \frac{\mu}{2}.$$

Differentiate with respect to μ; then

$$\cos\phi \frac{d\phi}{d\mu} = \frac{1}{2} \quad \text{or} \quad d\phi = \frac{1}{2\cos\phi} d\mu.$$

If now $\mu_2 - \mu_1$ be substituted for $d\mu$, $d\phi$ gives the separation in radian measure. Thus

$$d\phi = \frac{1}{2\sqrt{1-\frac{1}{4}\mu^2}} 7{\cdot}1\ 10^{-5} = 5{\cdot}59\ 10^{-5} \text{ in rads.}$$
$$= 5{\cdot}59\ 10^{-5} \times 57 \times 60 \times 60 = 11{\cdot}5 \text{ in secs.},$$

which would not be visible with an ordinary spectrometer.

The Analysis of Polarised Light. Let us suppose that as a result of some experiment we have a beam of light and that it is desired to ascertain its nature, whether it is plane polarised or circularly polarised or elliptically polarised, and, if so, in what directions the axes of the ellipse lie and what is the ratio of their lengths. We require for this purpose an analysing nicol and either a Babinet's compensator or a quarter wave plate.

Babinet's compensator in its simplest form consists of two quartz wedges of equal angle, one of which can be moved past the other by means of a micrometer screw, so that together they form a parallel plate of variable thickness. The fixed wedge A is cut with its axis in the plane of the paper in the direction of the arrow, the movable wedge B is cut with its axis perpendicular to the plane of the paper. Of course in the figure the angles of the wedges are much exaggerated. If a parallel beam falls normally on the compensator and enters the wedge A, it is decomposed into two beams, one polarised in the plane of the paper and one polarised at right angles to the plane of the paper. The beam polarised in the plane of the paper travels faster in the wedge A than the other and consequently suffers a phase retardation with reference to that other. In the wedge B it travels slower than the other beam and consequently its phase is accelerated with reference to the other. At the point where the thicknesses of both wedges are the same, the acceleration annuls the retardation and the two components leave the compensator in the same phase. On both sides of this point there is a phase difference proportional to the distance from the point.

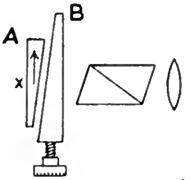


FIG. 208.

In front of the fixed wedge cross-wires are fixed, and an eyepiece focusses on these through the analysing nicol.

Before using the compensator the readings of the micrometer screw must first be evaluated in terms of the wave-length of the light in question. This is done by allowing light which is plane polarised in a plane making approximately an angle of 45° with the axes of the two wedges to enter the instrument. Where the phase difference produced by the two wedges is zero or a multiple of 2π, the light is plane polarised in the same plane after passing through the wedges and it is polarised in this plane nowhere else. Consequently, if the compensator be removed, the analysing nicol set to extinguish the incident light and the compensator again inserted, there are black bands parallel to the edge of the wedge where the phase difference is a multiple of 2π. The movable wedge is adjusted with the micrometer screw until one of these bands is on the cross-wires and the reading noted; the screw is then turned until the next band is on the cross-wires and the reading again noted. The difference corresponds to a phase difference of one wave-length.

In order to determine which of the black bands corresponds to the phase difference zero, i.e. the point where the quartzes are equally thick, it is only necessary to illuminate the compensator with white light. Then this band alone appears black; the positions of the others are different for the different components of white light and consequently they appear coloured.

In investigating the state of polarisation of a beam of light the first step is to see whether it can be extinguished with the analyser alone. If so, it is plane polarised and its plane of polarisation is parallel to the principal section of the analyser, i.e. to the shorter diagonal of its end face.

If the beam cannot be extinguished with the analyser alone, set the compensator so that it produces a phase difference of $\lambda/4$ and rotate it about the line of vision until a dark band moves on to the cross-wire. Then the analyser must be adjusted to make this band as black as possible, i.e. to make the extinction perfect at the cross-wires. From consideration of (39) it is evident that the principal sections of the two quartz wedges now give the directions of the axes of the elliptic vibration. Also if the principal section OA of the analyser (fig. 209) makes an angle θ with OB, the principal section of one of the quartzes, the vibration after leaving the compensator is in the direction BC—if we make the usual assumption that the vibration takes place at right angles to the plane in which it is polarised—and consequently the axes of the ellipse are in the ratio OB to OC. Or, in other words, if the principal section of the analyser makes an angle θ with the principal section OB of one of the wedges, then

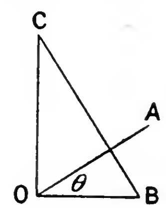


FIG. 209.

$$\frac{\text{length of axis of ellipse parallel to OB}}{\text{length of axis perpendicular to OB}} = \cot\theta.$$

If $\theta = 45°$ the light is circularly polarised. It is easy to tell when the principal section of the analyser is parallel to OB or OC because then the interference bands disappear.

The quarter wave plate, as its name implies, is a plane parallel plate cut from a crystal, for example, mica or quartz, and of such a thickness that the ordinary and extraordinary beams suffer a relative phase difference of $\lambda/4$ in passing perpendicularly through it. It can be used in the same way as the Babinet compensator, but, unlike the latter, is available only for one wave-length, usually for sodium light. If another colour is to be investigated, the thickness of the plate must be different.

Production of Elliptically Polarised Light. By means of a nicol and a Babinet's compensator any kind of elliptically polarised light may be produced from natural light. For if the nicol is set with its principal section at 45° to the principal sections of the compensator, and if the light passes first through the nicol and then through the compensator, the two beams into which it is divided by the latter have equal amplitudes, and vibrate at right angles to one another, and their phase difference varies according to the part of the compensator passed through. If the y and z axes be taken in the principal sections of the compensator, the two beams may be represented in this case by

$$\eta = a \sin \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right), \quad \zeta = a \sin\left\{\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) + \delta\right\},$$

or if we make the point under observation the origin of coordinates, by

$$\eta = a \sin \frac{2\pi t}{\tau}, \quad \zeta = a \sin\left\{\frac{2\pi t}{\tau} + \delta\right\},$$

where δ is the phase difference produced by the compensator. If $\delta = 0$, this gives a plane polarised wave, if $\delta = \pi$, it gives a plane polarised wave polarised in a plane at right angles to the first; if $\delta = +\pi/2$, it gives a right-handed circularly polarised wave, if $\delta = 3\pi/2$, it gives a left-handed circularly polarised wave. If δ lies between 0 and π, consider the time when $t = 0$; then $\eta = 0$, $\zeta = a \sin \delta$, the particle is above the middle of its range and its velocity is positive. Consequently the ellipse is right-handed. If δ lies between π and 2π, ζ is negative, the particle is below the middle of its range, the velocity is still positive, and consequently the ellipse is left-handed. Fig. 210 shows how the nature of the vibration varies with δ.

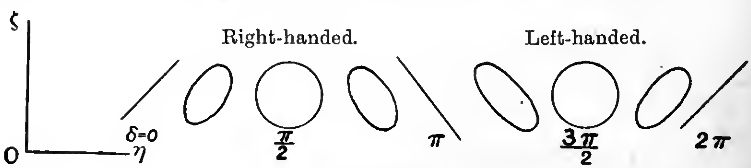


FIG. 210.

Partially Polarised Light. If a beam of light contains natural light in addition to plane polarised, circularly polarised, or elliptically polarised light, then it cannot be extinguished with a Babinet's compensator and analyser. A minimum of intensity can however be obtained and thus a rough idea formed of the quantity of natural light present.

The presence of a little plane polarised light in a beam of natural light can be detected very readily by means of the Savart polariscope.

This is made of a thin plate of quartz which has its faces at an angle of 45° to the axis; this plate of quartz is cut into two, and the two pieces placed on the top of one another with their principal sections at right angles and cemented together in this position. They are then mounted in a tube in front of a nicol prism with the bisector of the angle between their principal sections parallel to the principal section of the nicol.

If the polariscope is turned to view a source of plane polarised light, we have the ordinary case of interference produced by convergent light in a crystal plate between nicols. The interference figures are straight fringes parallel to the bisector of the angle between the principal sections. They become most distinct when the plane of polarisation of the incident light is parallel to the bisector. If the incident light is white, they are, of course, coloured.

If the incident light contains natural light in addition to plane polarised light, a uniform illumination is superimposed on the interference fringes and the latter are consequently dimmer. But, even when only a small percentage of the incident light is plane polarised, they are sufficiently distinct to detect its presence. Since they are sharpest when parallel to the plane of polarisation of the light, the direction of the latter can readily be determined.

The light of the sky is partially plane polarised and easily gives fringes with the Savart polariscope.

Fresnel's Rhomb. Fresnel's rhomb is a substitute for the quarter wave plate. It is a parallelipiped of crown glass; on one end of it a beam of light falls normally, is twice reflected internally at an angle of incidence of 55°, and then emerges normally from the other end. If the incident light is plane polarised, the emergent light consists of a beam polarised in the plane of the figure and a beam polarised at right angles to the plane of the figure, the relative phase difference being one quarter wave-length. Thus if the incident light is polarised in a plane making an angle of 45° with the plane of the figure, the emergent light is circularly polarised. Fresnel arrived at this result first on theoretical grounds and constructed the rhomb to test it. The rhomb has the advantage over the quarter wave plate that it produces approximately a phase difference of $\lambda/4$ for all the components of white light although strictly accurate only for one colour.

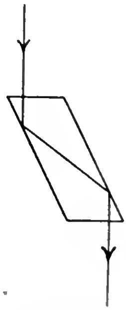


FIG. 211.

Double Refraction Produced by Strain. If a block of glass between two crossed nicols is subjected to strain, the field lights up, showing that the glass becomes double refracting during the strain, then, when the strain is removed, it becomes dark again, provided of course that the glass has not been overstrained. This is a well-known laboratory experiment; the piece of glass is placed in a metal frame such as is shown in fig. 212, and the strain applied by turning a screw. Examination between crossed nicols is also a means of testing the quality of optical glass. If the glass has not been properly annealed, if it has cooled too rapidly and has thus become strained, it will not show a dark field. Glasses which have been purposely cooled rapidly, give

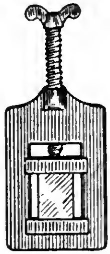


FIG. 212.

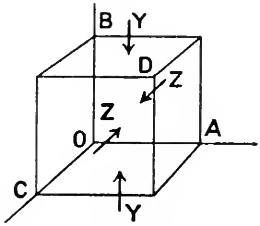


FIG. 213.

very interesting interference figures, and usually form part of a laboratory outfit for polarisation experiments.

Let the block of glass ABC represented in fig. 213 be compressed by a normal pressure of Y lbs./sq. inch on the faces AC and BD, and let it be compressed by a normal pressure of Z lbs./sq. inch on the faces AB and CD. This includes the case of tension on these faces if the signs of Y and Z are understood algebraically. Then, if a wave is incident perpendicularly on the face BC, it is decomposed on entering the block into two waves which are polarised in the planes AC and AB, and after passing through the block the component waves have a relative phase difference given by

$$\delta = c\,(Y - Z)l, \quad . \quad . \quad . \quad . \quad (41)$$

where l is the thickness of the block, and c is a constant which can be obtained by experiment; δ is expressed in radians.

The Optical Determination of Stress. Prof. E. G. Coker has recently applied the formula (41) very successfully to the determination of stress in engineering problems. Suppose, for example, it is desired to find what the stress is at the different points in a beam in a certain structure, so as to be able to leave a sufficient margin of safety. Elementary calculations give only the average stress, and the accurate mathematical theory can be applied only to the simplest cases and breaks down if we wish, for example, to consider the effect of a notch or a hole on the distribution of stress. Brewster had suggested making a glass model of the piece to be investigated and examining it in polarised light under strain, but glass does not easily give a measurable effect. Prof. Coker has substituted xylonite for glass with the best results. It is much more compressible than glass, has a Young's modulus of 3 10^5 lbs./sq. inch, and is easily worked.

In one of the experimental arrangements used the source of light is a row of incandescent lamps behind a diffusing screen. The rays from the screen are plane polarised by reflection at a black mirror. This takes the place of the polarising nicol. They then pass through the model and through an analysing prism to the eye of the observer. The analyser is set to extinction. The parts of the model which are unstrained appear black, and the other parts appear tinted with different colours according to the degree of double refraction and strain they experience. By having in the field beside the model under investigation a comparison model of a simple form, for which the distribution of stress is thoroughly understood, and by comparing tints, the difference of stresses at any point in the unknown model can be evaluated from its colour directly in lbs./sq. inch. In the case of simple tension or compression where either Y or Z is zero, the absolute value of the stress can be directly determined.

The principle of the method is thus the interference between crossed nicols in parallel light, and the intensity is given by formula (19), namely

$$I = \sin^2 2a \sin^2 \frac{\delta}{2} \quad . \quad . \quad . \quad . \quad (19)$$

If the directions of the stresses are such that $a = 0$, the field is dark, the colours cut out, and only the direction, not the magnitude of the effect, can be determined. To obviate this difficulty Prof. S. P. Thompson has inserted two quarter wave plates, one between the polariser and the model, and the other between the analyser and the model. Both quarter wave plates are set with their axes at 45° to the principal sections of the polariser and analyser.

After passing through the quarter wave plate the light is circularly polarised, but on entering the xylonite plate it resolves into two plane polarised waves of equal amplitude. If the amplitude before entering the quarter wave plate is taken as unity, the amplitude of each of these plane polarised vibrations is $1/\sqrt{2}$. They have a relative phase difference of $\frac{\pi}{2}$ on entering the xylonite and a relative phase difference of $\frac{\pi}{2} + \delta$ on leaving it. They may be represented by the lines OX and OY in fig. 214, where the phase of each vibration is marked on it. Now let OQ be the direction of the axis of the second quarter wave plate and OP a line at right angles to it. Then the vibration OX resolves into OQ and QX, and the vibration OY into OP and PY. The phases of these vibrations are marked on them on the assumption that the vibrations parallel to OQ experience an acceleration of phase $\frac{\pi}{2}$ in passing through the quarter wave plate.

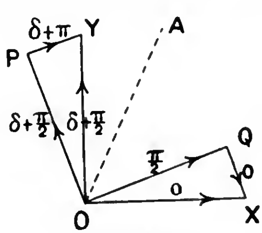


FIG. 214.

Let OA be the direction of the principal plane of the analyser; it makes, of course, an angle of 45° with OQ. Then OQ and QX resolved along OA give amplitudes OQ cos 45° and $-$ QX cos 45° with phases $\frac{\pi}{2}$ and 0. These compound into a single amplitude

$$\sqrt{(\text{OQ}\cos 45°)^2 + (\text{QX}\cos 45°)^2} = \tfrac{1}{2}$$

with a phase $= \frac{\pi}{2} - \angle\text{QOX}$. Similarly OP and PY resolved along OA give amplitudes OP cos 45° and PY cos 45° with phases $\frac{\pi}{2} + \delta$ and $\pi + \delta$ which compound into a single amplitude

$$\sqrt{(\text{OP}\cos 45°)^2 + (\text{PY}\cos 45°)^2} = \tfrac{1}{2}$$

with a phase $= \frac{\pi}{2} + \delta - \angle\text{QOX}$. The light issuing from the analyser consists therefore of two waves of equal amplitude $\frac{1}{2}$ with a relative phase difference δ. These give an intensity of

$$(\tfrac{1}{2})^2 + (\tfrac{1}{2})^2 + 2(\tfrac{1}{2} \times \tfrac{1}{2})\cos\delta$$
$$= \cos^2\frac{\delta}{2},$$

no matter what the orientations of the principal directions of strain are with reference to the nicols. Thus the $\sin^2 2a$ factor in (19) has disappeared. The achromatic black cross usually superimposed on the colours is therefore removed.

In order to have as large a field as possible Prof. Thompson has made quarter wave plates up to 20 inches in diameter. This is a record for size.

The Kerr Effects. A transparent dielectric such as glass, olive oil, carbon bisulphide, or turpentine becomes double refracting when placed in a strong magnetic field. This was first discovered by Dr. Kerr of Glasgow in 1875. He experimented with a block of glass in the opposite ends of which holes were drilled; in these holes were fixed the terminals of the secondary of an induction coil. Since the break of the primary in an induction coil is much sharper than the make, there was thus a strong intermittent electric field of constant direction between the terminals. The terminals were connected to a micrometer spark gap across which a constant stream of sparks passed. By adjusting the spark gap the potential difference between the terminals could be regulated. No sparks, of course, passed between the terminals directly.

The thickness of glass between the terminals was about one-quarter inch, and through this space at right angles to the lines of force was passed a beam of plane polarised light. Before the induction coil was started this beam was extinguished by an analysing nicol, but on starting the coil the field lighted up, the full effect taking 20 or 30 seconds to develop, and the illumination could not be extinguished again by rotating the analyser. The effect was most marked when the plane of polarisation of the incident light made an angle of 45° with the lines of electric force; when it was parallel or perpendicular to them there was little or no effect. Thus, under the influence of the electric field, the glass becomes double refracting and polarises the light in planes parallel and perpendicular to the lines of force. The effect is independent of the direction of the field, and its magnitude is proportional to the square of the field strength.

In general, if a plane polarised beam of light is incident on a metal mirror, the reflected beam is elliptically polarised, but if the incident beam is polarised either in or at right angles to the plane of incidence, the reflected beam is plane polarised in the same plane. In 1877 Kerr discovered that, when a beam of light polarised in or at right angles to the plane of incidence is reflected from the magnetised polished pole piece of a powerful electromagnet, as a result of the reflection its plane of polarisation is rotated through a small angle.

EXAMPLES.

(1) Calculate the thickness of a quarter wave plate of quartz for sodium light.

(2) Two simple harmonic vibrations in directions at right angles to one another compound into a parabolic vibration, if the period of the one is double the period of the other, and if they both start from the extremities of their ranges simultaneously.

(3) Any number of simple harmonic vibrations in different directions, differing in phase but having the same periodic time, compound into an elliptic vibration.

(4) The components of an elliptic vibration are given by $x = a \sin \omega t$, $y = b \sin (\omega t + \delta)$. Show that the directions of the axes of the ellipse are obtained from

$$\tan 2\phi = \frac{2ab}{a^2 - b^2} \cos \delta,$$

where $\tan \phi = y/x$.

(5) On the assumption that the eye can always match the intensities of the two halves of the field in a half-shadow apparatus to 1 per cent., find how accurately the rotation can be measured with a half-shadow apparatus, when the planes of polarisation of the two halves of the field are inclined to one another at 7°, 5°, and 3°. Of course, if the angle is too small, both halves of the field are dark when the intensity is matched, and the setting can no longer be made to 1 per cent. of the intensity.

(6) In front of the slit of a spectroscope a student places two nicols with a quartz plate between them. The nicols are mounted so that they can rotate about the axis of the collimator and are set with their planes crossed. The quartz plate is cut with its surfaces at right angles to the optic axis, and is set with its optic axis parallel to the axis of the collimator. The arrangement is directed towards an incandescent mantle and the student observes that the spectrum is crossed by dark bands. How are these bands caused and where are they situated, given that the quartz is 2 cms. thick? (The dispersion of the optical rotation of quartz is given in the tables at the end of the book.)

(7) A student makes up four solutions of sugar containing respectively 30·5 gms., 22·76 gms., 20·4 gms., and 17·53 gms. in the 100 c.cs. of solution.

He puts them in succession in a simple saccharimeter consisting of a tube between two nicols. The source of light is a sodium flame and the tube is 23·5 cms. long. He finds the rotations produced to be 49·5°, 36·1°, 30·5°, and 26·8° respectively. What is his result for the specific rotation of sugar?

(8) Three plates of left-handed quartz cut perpendicularly to the axis and of thickness 4·73 mm., 5·96 mm., and 9·05 mm. give apparent rotations of 77·5°, 49°, and − 18·3° when placed between nicols and used with a sodium flame. What result do these observations give for the specific rotation of quartz?

(9) The tube of a Soleil saccharimeter is 20 cms. long, and when it is filled with a certain solution of cane sugar the wedge has to be displaced 98·2 divisions on the scale to make the two halves of the field equally bright again. The solution was formed by dissolving 10·07 gms. up to a bulk of 41·0 c.cs. If the specific rotation of cane sugar be assumed, calculate the value of a scale division in angular measure. The correct value for the instrument should be ·345°.

(10) An aqueous solution of cane sugar, which contains other optically active substances, is contained in a tube 20 cms. long and gives a rotation of 50·2°. The source of light is a sodium flame. Fifty c.cs. of the solution are then measured off, 5 c.cs. of strong hydrochloric acid added, the mixture kept at a temperature of 70° C., for ten minutes, cooled to 20° C., and poured back into the same tube. The rotation is now found to be − 15·2° C. What was the original concentration of the cane sugar in gms. per 100 c.cs. of solution?

(11) A piece of flint glass is placed between the poles of an electro-magnet, the pole pieces of which are bored so that a beam of plane polarised light can traverse the glass in the direction of the lines of force. The distance traversed in the glass is 2 cms., and the average field strength is 5000 c.g.s. units. How much is the plane of polarisation rotated?

PART III.

SPECTROSCOPY AND PHOTOMETRY.

CHAPTER XIV.

SPECTROSCOPY: EARLIER WORK.

If a beam of white light falls upon a glass prism, it is changed by the prism into a series of coloured beams—red, orange, yellow, green, blue, indigo, violet—each of which is deviated to a different extent. The deviation of the red beam is least and of the violet beam is greatest, and the other colours come in between in the order stated. This experiment was first studied by Newton. In his arrangement a beam of sunlight entered a darkened room through a round hole in the shutter, then passed through a glass prism, and was finally received on a screen. Owing to the different deviations of the different colours a coloured band was produced on the screen, the least deviated end being red and the other colours coming in the above order. This band of colours was called by Newton a spectrum.

Before Newton's time the different colours were supposed to be made out of the white light by the prism during the passage of the light through it. Newton allowed the beam of sunlight to pass in succession through two prisms which were crossed, that is, had their refracting edges at right angles to one another, and found a single straight spectrum still produced.

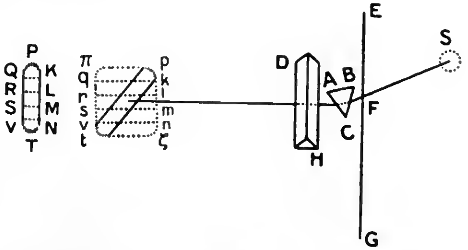


Fig. 215.

Fig. 215, which is taken from his "Opticks," Bk. I, shows his arrangement. S is the sun. PT was the spectrum produced when only the first prism was used, *pt* the spectrum when both were used. Since any portion of the first spectrum, SMVN for example, was not altered any further by passing through the second prism, i.e. did not give rise to a horizontal spectrum *smnv* but preserved its own character only suffering additional deviation, Newton came to the conclusion that the prism did not change the nature of the light that passed through it. White light was due merely to the superposition of a great number of coloured beams. The prism only separated these beams out. It did not produce them. They existed in the white beam before it reached the prism. The different deviation was due to each coloured beam having its own index of refraction.

Newton's views as to the constitution of white light prevailed until recently, but now the first prism is supposed to make the colours. This is discussed fully in Chapter XXI.

Newton worked with a circular aperture. Thus his spectrum consisted of a series of blurred images of the sun formed by the successive colours somewhat after the manner of the pinhole camera. All these images overlapped considerably and consequently the spectrum was not pure. Wollaston in 1802 used instead of the circular aperture a narrow slit, which was parallel to the refracting edge of the prism, and observed that in the purer spectrum obtained in this way there were a number of black bands. He did not follow the matter up, and it was left for Fraunhofer to discover their invariable position in the spectrum and consequent great importance for scientific measurement.

Fraunhofer was engaged at Munich in a glass works in the manufacture of telescope lenses. He was the first to use a convex lens to make the light from the slit parallel before it fell on the prism and to use a telescope to examine the spectrum, the same arrangement that we employ to-day. By using different prisms he was able to show that the black lines have fixed positions in the solar spectrum, and he recognised them as radiations in which the solar spectrum is deficient. He mapped their positions, denoted the principal ones by letters of the alphabet, and used them as standard radiations for the determination of indices of refraction. His notation is still used and has been extended since his time beyond the limits of the visible spectrum. The wave-lengths and positions of the principal Fraunhofer lines in

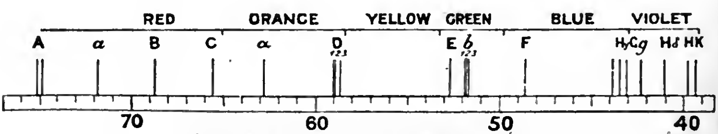


FIG. 216 (from "Watson's Physics").

the normal spectrum are shown in fig. 216. The wave-lengths are also given in the tables at the end of the book.

Fraunhofer was the first to use the sodium flame as a monochromatic light source. He also made diffraction gratings, both by stretching fine wires between two screws with equal threads and by ruling lines with a diamond point on glass, and used these gratings to make accurate determinations of the wave-length of sodium light and of the

Fraunhofer lines. Fraunhofer's paper on the determination of wave-lengths by the diffraction grating was published in 1821.

§ An instrument used for examining spectra is called a spectroscope or spectrometer. There are many different kinds of spectroscope according to the special purpose for which they are to be used. This section deals only with the simpler ones for use with the visible spectrum.

Any spectrometer designed for measuring angles can also be used for examining spectra, but if it is to be used only for the latter purpose its construction may be simplified. First of all the prism or diffraction grating is fixed on the prism table. Then, as the spectrum always appears at the same place on the circle, the latter does not require to be divided the whole way round. It is enough if it is divided through an arc of 60° or 90°. Also it is not necessary for the scale to be divided in degrees; any arbitrary unit is sufficient. Finally the prism is usually covered in to prevent stray light entering the telescope; if the spectrum is at all faint, it will be otherwise swamped by this stray light.

The performance of a spectroscope is measured by its resolving power and the brightness of the spectra it produces. The brightness of the spectrum depends in the first place on the aperture of the instrument, i.e. the ratio of the diameter of the collimator object glass to its focal length. The aperture is usually about $\frac{1}{12}$ and rarely exceeds $\frac{1}{8}$, because, unless the lenses are specially corrected, what is then gained in brightness is lost in definition. Then the size and material of the prism are important. The width of the beam that emerges from the collimator object glass is usually too great for the prism, and consequently some of the light passes the edge of the latter and misses it altogether. Also some of the light is lost by external reflection on entering the prism and by internal reflection on leaving it. In the case of a 60° prism at minimum deviation from 10 to 20 per cent. of the incident light is lost in this way, the percentage lost increasing with the index of refraction of the glass. Light is also lost by absorption in the glass of the prism itself. It is difficult to give figures for this loss, as the fraction transmitted diminishes with the length of the path in the glass and is also much less for the violet end of the spectrum, but in the case of a 60° dense flint glass prism, the length of side of which is one inch, probably one half of the blue light incident on the prism is absorbed. Less is absorbed in the case of a crown glass prism.

In spectra lines occur often close together; for example, the yellow sodium line is double, consisting of two components the wave-lengths of which are 5·890 and 5·896 10^{-5} cm. Let us suppose that λ and $\lambda + d\lambda$ are the wave-lengths of two such lines, and that, when the slit is made as narrow as possible, the two lines can be seen just resolved; then $\lambda/d\lambda$ is said to be the resolving power of the prism for that part of the spectrum. It has been shown in Chapter X that if the incident beam fills the face of the prism the theoretical value of the resolving power is equal to $t\,\frac{d\mu}{d\lambda}$, where t is the length of the base of the prism and μ the index of refraction of its material. Of course this formula assumes perfect accuracy of all the optical surfaces, and the theoretical resolving power is not attained in cheap spectroscopes. Since $\frac{d\mu}{d\lambda}$ is greater for flint glass than for crown glass, the resolving power of a flint glass prism is greater than that of a crown glass prism of the same shape and size.

The resolving power of a spectroscope specifies the amount of detail visible in its spectrum. There is an allied quantity, the dispersion, which specifies the length of the spectrum, and of which there is no universal definition. For example, it may be said of a grating spectroscope that with a certain eyepiece it gives an apparent dispersion of about 30°, the whole of the spectrum being in the field of view at once. This means that the two ends of the visible spectrum appear to subtend an angle of 30° at the eye, and obviously in this sense the dispersion depends on the eyepiece used. It is more usual, however, to make the definition independent of the eyepiece and to specify the dispersion for any one position of the prism by $\frac{d\theta}{d\lambda}$, where θ is the reading of the position of the telescope on the circle. The dispersion in this sense varies with the position of the prism; it is easily found experimentally to be a minimum at minimum deviation, for if we turn through minimum deviation the spectrum is shortest there.

Let us assume that the prism is set at minimum deviation for a wave-length near the middle of the spectrum, say the sodium lines, and that μ is the index of refraction and θ the deviation for that wave-length. Let A be the refracting angle of the prism. Then

$$\mu = \frac{\sin\frac{A+\theta}{2}}{\sin\frac{A}{2}}.$$

If we differentiate both sides of this equation with respect to μ,

$$1 = \frac{\cos\frac{A+\theta}{2}}{\sin\frac{A}{2}}\,\frac{1}{2}\,\frac{d\theta}{d\mu} \quad \text{or} \quad \frac{d\theta}{d\mu} = \frac{2\sin\frac{A}{2}}{\cos\frac{A+\theta}{2}}.$$

Hence

$$\frac{d}{d\lambda} = \frac{d\theta}{d\mu}\,\frac{d\mu}{d\lambda} = \frac{2\frac{d\mu}{d\lambda}\sin\frac{A}{2}}{\cos\frac{A+\theta}{2}} = \frac{2\frac{d\mu}{d\lambda}\sin\frac{A}{2}}{\left(1-\sin^2\frac{A+\theta}{2}\right)^{\frac{1}{2}}} = \frac{2\frac{d\mu}{d\lambda}\sin\frac{A}{2}}{\left(1-\mu^2\sin^2\frac{A}{2}\right)^{\frac{1}{2}}}.$$

For a 60° prism this becomes

$$\frac{d\theta}{d\lambda} = \frac{\frac{d\mu}{d\lambda}}{\left(1 - \frac{\mu^2}{4}\right)^{\frac{1}{2}}}.$$

If one line of the sodium doublet is adjusted for minimum deviation, the cross-wires set on it, and the reading taken, and then the prism adjusted so that the other line of the doublet is at minimum deviation, the cross-wires set on it, and the reading taken, the difference of the two readings gives $d\theta$ for the difference of wave-length of the two lines. Since, however, the position of a line is stationary at minimum deviation, $d\theta$ has the same value if the prism is kept in the same position for both readings. Hence the above formula gives the dispersion, according to our definition, at any point in the spectrum.

For rough work it can be used for the average dispersion throughout the whole spectrum. For example, suppose it is desired to compare the dispersion of 60° prisms made respectively of the crown and flint glasses given on p. 61. Calculate $\frac{d\mu}{d\lambda}$ by taking the values for the wave-lengths C and F. Then for the crown glass

$$\frac{d\mu}{d\lambda} = \frac{\cdot 00856}{1{\cdot}701\ 10^{-5}} = 503$$

and for the flint glass

$$\frac{d\mu}{d\lambda} = \frac{\cdot 01722}{1{\cdot}701\ 10^{-5}} = 1012.$$

Hence for the crown glass

$$\frac{d\theta}{d\lambda} = \frac{503}{\left(1 - \frac{1{\cdot}517^2}{4}\right)^{\frac{1}{2}}} = 7{\cdot}7\ 10^2$$

and for the flint glass

$$\frac{d\theta}{d\lambda} = \frac{1012}{\left(1 - \frac{1{\cdot}621^2}{4}\right)^{\frac{1}{2}}} = 17{\cdot}2\ 10^2;$$

thus the dispersion of the flint glass is more than twice the dispersion of the crown glass.

Direct Vision Spectroscopes. In a spectroscope made with a single prism the light is deviated as well as dispersed. Hence if we desire to examine the spectrum of a flame or of the sun we do not look in the direction of the source; allowance must be made for the deviation produced by the prism. This can easily be done in the case of large instruments mounted on stands and resting on a table, but it is much more difficult in the case of light instruments intended to be held in the hand and carried about in the pocket. Consequently such instruments are usually made so that the middle part of the spectrum is not deviated at all, so that the eye looks in the direction of the light to be examined. This can be done in two ways, either by using a series of prisms made of two kinds of glass or by using a grating mounted on a single prism. In fig. 217, which represents the first method, the

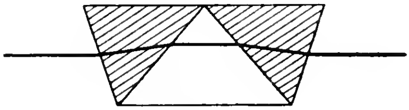


FIG. 217.

middle prism is made of very dense flint glass and the two outside prisms of crown glass. The angles of the prisms are chosen so that the middle part of the spectrum passes through undeviated as illustrated by the ray shown. The dispersion of the middle prism is much more than twice as great as that of the outside prisms. Consequently they are not able to neutralise it, and the system produces a spectrum. This arrangement was first used by Amici in 1860. Five prisms are sometimes used instead of three, the second and fourth being of flint and the others of crown glass. The prisms are cemented together with Canada balsam. The refracting angle of the flint prism is usually so great that if there were an air film between the prisms the light would not be able to enter it, but would be totally reflected back into the crown glass.

Fig. 218 illustrates the second method. ABC is a prism, on the face AB of which a contact copy of a diffraction grating is mounted. White light falls perpendicularly on the face AC, enters the prism undeviated, then falls on the grating and is diffracted out in all directions. The wave-length diffracted out in the direction PQ at an angle θ with PN, the normal to AB, is given by

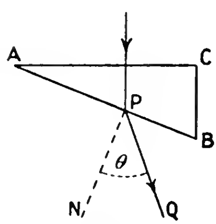


FIG. 218.

$$\lambda = e \sin \theta,$$

where e is the distance between the two adjacent rulings. If $\theta = \angle BAC$ the light of wave-length λ is undeviated. Hence in order to produce an undeviated spectrum it is only necessary to make $\angle BAC$ equal to the value of θ for the middle of the spectrum. If there are 14,000 lines to the inch, this will be about 20°.

Pocket direct vision spectroscopes have usually a collimator but no telescope. The eye comes close up to the prism, and nothing comes between the eye and the prism. There is thus usually no means of measuring the wave-length of any line, though in some cases this can be done by a scale which is seen by reflection in the side of the prism next the eye.

Wave Length Spectroscopes. In some instruments the wave-lengths of the spectral lines can be read off directly; figs. 219 and 220 represent well-known examples of this kind.

In fig. 219 A is the slit, B the collimator lens which renders the light parallel, C a plane transmission grating which is attached to the collimator, D the telescope object glass, E the cross-wires, and F

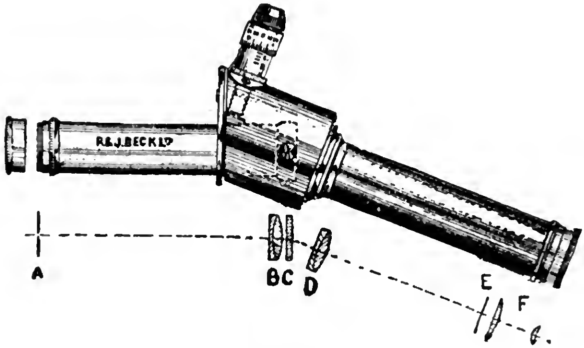


FIG. 219.

the eyepiece. The telescope is hinged on the collimator and is rotated by a micrometer screw. When the cross-wires are brought into coincidence with a spectral line the reading of the drum gives the wave-length of the line.

The spectroscope represented in fig. 220 is of the constant deviation

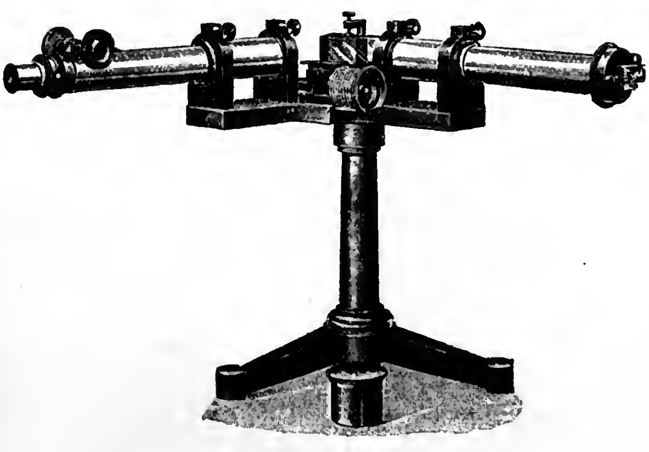


FIG. 220.

or fixed arm type. The telescope and collimator remain fixed at right angles to one another, and when the prism table is rotated the spectrum passes through the field, and a drum passes under an index which gives the wave-length of each spectral line as it comes into coincidence with the cross-wires. Fig. 221 represents the prism, which is made in one piece, but can be regarded as built up of two 30° prisms and a right-

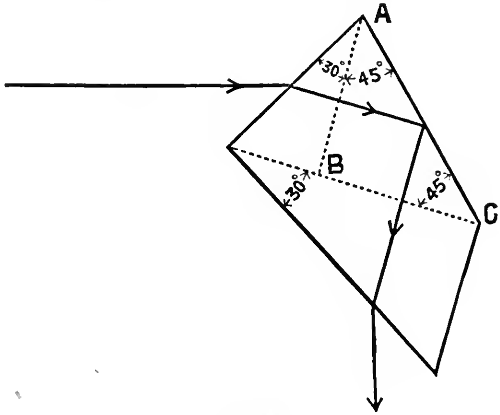


FIG. 221.

angled prism. As the prism rotates, every ray in the spectrum is incident in succession at 45° on the face AC and then makes equal angles with the surface on entering and leaving the prism. It is consequently turned through exactly 90° by the prism. The action of the prism on it is otherwise that of a 60° prism at minimum deviation. Each ray of the spectrum comes into minimum deviation as it passes the cross-wires.

In addition to the ordinary model this instrument is made with prism and lenses in quartz instead of glass, and the drum calibrated for the ultra-violet, and also with a quartz prism and concave mirrors instead of lenses, and the drum calibrated for the infra-red.

It is very suitable for use as a monochromatic illuminator, i.e. as a means of illuminating a given surface or aperture with the different colours of the spectrum in succession. For this purpose it is necessary only to remove the eyepiece, put a diaphragm in its place, and then point the telescope at the surface or aperture in question.

Calibration of a Spectroscope. In the general case, however, the scale gives the position of the cross-wires only in arbitrary units, and then before using the instrument for mapping unknown spectra it is necessary to calibrate it, i.e. evaluate its scale in terms of wave-lengths.

The way to do this is to read the positions of lines of known wave-length. For this purpose the following lines are useful :—*

* The method of producing these lines is described on pp. 246-251. There is also a table of wave-lengths at the end of the book.

Source.	*Wave-length.*	*Colour.*
Sodium bicarbonate on a platinum wire in bunsen	5890·0 *	Orange
	5895·9	
Thallium chloride on a platinum wire in bunsen	5350·7	Green
Lithium sulphate † on platinum wire or asbestos in bunsen	6707·8	Red
Potassium nitrate on platinum wire in bunsen .	7668 } 7702 }	Red
	4044 } 4047 }	Violet
Hydrogen vacuum tube	α 6563	Red
	β 4861	Greenish-blue
	γ 4340	Violet
	δ 4102	
Mercury vacuum tube or mercury arc ‡. . . .	5769·6 } 5790·7 }	Yellow
	5460·7	Green
	4358·2	Violet
Cadmium spark in air	6438·5	Red
	5058·8	Green
	4799·9	Blue
Zinc spark in air	6363·7	Red
	4912·0 } 4810·5 } 4722·1 }	Blue

When the telescope cross-wires have been set in succession on some of these lines and the scale-readings taken, the latter are plotted against the wave-lengths and a continuous curve drawn through the points. This curve is known as the calibration curve of the spectroscope. In order to determine the wave-length of an unknown line, it is only necessary to set the cross-wires on it and take the scale reading. The wave-length corresponding to the latter can at once be read off from the curve.

Sometimes instead of plotting the scale reading against the wave-length λ, it is plotted against the oscillation frequency or the number of wave-lengths in a centimetre $1/\lambda$, or even against $1/\lambda^2$. In this way a straighter calibration curve is obtained, but of course it takes longer to get an unknown wave-length from it.

In most good spectroscopes the slit is symmetrical; when it is opened both jaws move equally in opposite directions, and consequently, if the cross-wires are set on the middle of a line, the reading is always the same, no matter how wide the slit is. In cheap instruments the slit is unsymmetrical; only one jaw moves and the other remains fixed. In this case the cross-wires must be set on the fixed edge of the image.

If the slit is made very narrow the spectrum is crossed by streaks

* These wave-lengths are expressed in Ångström units (A.U.), i.e. 10^{-8} cms. or 10^{-10} metres. This unit is often referred to in America as the "tenth metre". Wave-lengths are also often expressed in micromillimetres ($\mu\mu$), i.e. 10^{-7} cms., and microns (μ), i.e. 10^{-4} cms.

† Lithium chloride is often recommended but deliquesces and makes the bunsen and table sticky.

‡ Mercury has also lines at 6232, 6152, 4959·7, 4916·4, 4078·1, 4046·8, 3650, 3131, 3126.

at right angles to the spectral lines. This is due to dirt between the jaws; the dirt may be to some extent removed by opening the jaws and rubbing them gently with a piece of match.

The spectral lines are always curved, being convex towards the refracting edge of the prism. This is due to the fact that the rays do not all pass through the prism in a principal section, i.e. a section at right angles to the refracting edge. The central ray of the pencil from the middle of the slit passes through in a principal section, but the central rays of the other pencils are inclined to the principal section. They are thus incident on the face of the prism at a different angle, and also the refracting angle is virtually increased for them. They thus suffer a greater deviation.

Edser and Butler's Method of Calibrating a Spectroscope. If a thin parallel-sided film of air is enclosed between two plates of glass and light is allowed to pass through it perpendicularly, interference takes place between the light which passes directly through and the light reflected internally at both sides of the film, but owing to the latter being much fainter than the former the bands are too faint to be seen. If, however, both surfaces of the air film are covered with a thin film of silver that reflects about three-quarters of the light incident on it, then the intensities of the interfering beams are more nearly equal and the bands better defined. If e is the thickness of the air film, the condition for a black band is

$$2e = (n + \tfrac{1}{2})\lambda,$$

where n is any integer. If such a film is placed in front of the slit of a spectroscope, the spectrum is crossed by black bands wherever λ satisfies the above equation. Let us suppose that we know λ_1 and λ_2, the values of two of the wave-lengths at which black bands occur. Then

$$2e = (n_1 + \tfrac{1}{2})\lambda_1 \quad \text{or} \quad \frac{2e}{\lambda_1} = n_1 + \tfrac{1}{2}$$

and

$$2e = (n_2 + \tfrac{1}{2})\lambda_2 \quad \text{or} \quad \frac{2e}{\lambda_2} = n_2 + \tfrac{1}{2}.$$

These two equations give

$$2e\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right) = n_1 - n_2.$$

Similarly

$$2e\left(\frac{1}{\lambda} - \frac{1}{\lambda_2}\right) = n - n_2,$$

and finally on dividing the second of these equations by the first

$$\frac{\dfrac{1}{\lambda} - \dfrac{1}{\lambda_2}}{\dfrac{1}{\lambda_1} - \dfrac{1}{\lambda_2}} = \frac{n - n_2}{n_1 - n_2},$$

which gives

$$\frac{1}{\lambda}(n_1 - n_2) = \frac{1}{\lambda_2}(n_1 - n) + \frac{1}{\lambda_1}(n - n_2)$$

or

$$\frac{1}{\lambda} = \frac{n_1 - n}{n_1 - n_2}\frac{1}{\lambda_2} + \frac{n - n_2}{n_1 - n_2}\frac{1}{\lambda_1}.$$

Hence, since λ_1 and λ_2 are known and $n_1 - n_2$, $n - n_1$, and $n - n_2$ can be counted, λ can be calculated, and thus the whole spectrum calibrated.

To prepare the air film Edser and Butler recommend placing a little soft wax round the edges of the plates and pressing them together while looking through at a distant light. A series of images will then be seen by multiple reflection. When the surfaces are parallel, these images contract to a single image.

Hartmann's Dispersion Formula. Sometimes, instead of using a calibration curve for a spectroscope, it is more advantageous to calculate a table giving the wave-length in terms of scale readings. For this purpose the following empirical formula given by Hartmann is very suitable :—

$$\lambda = \lambda_0 + \frac{c}{s - s_0}.$$

λ is the wave-length, s the scale reading, λ_0, c and s_0 constants to be determined from the observations. The simplest way of determining them is to take three observed points on the curve, one near the middle and the other two, if possible, about a sixth from each end, and substitute the values of λ and s for these points in the equation. We thus obtain three equations for the three unknowns. When the constants have been determined the formula can be tested by means of the other observed points, and if it gives them satisfactorily, it can be used to draw up a table giving, for example, the value of λ for every 5′ increase in s.

Hartmann has also shown that the formula

$$\mu = \mu_0 + \frac{c}{(\lambda - \lambda_0)^{1\cdot 2}}$$

represents the index of refraction of optical glasses in the visible spectrum with a very high degree of accuracy. μ is, of course, the index of refraction, and μ_0, c and λ_0 are constants. This formula like the other one is purely empirical.

Care should be taken to verify a calibration curve or table at frequent intervals by taking the reading for the sodium lines, for it often happens that the collimator or prism table gets a slight knock which alters all the readings. In this case the calibration curve can usually be corrected by adding a small constant quantity to, or subtracting one from, all the scale readings.

The Production of Spectra. The spectrum used in the earlier part of last century was almost always the solar spectrum. We can, of course, get brighter light from the sun than from any terrestrial source, and at that time the electric arc and spark had not been developed. The disadvantage of the sun as a source, when it is there, is its motion in the heavens. With fixed apparatus it is always moving off the slit. So an instrument called a heliostat was used. This consists of a mirror driven by clockwork and placed outside the window of the room, which directs a beam of light into the spectroscope and which by its motion compensates the motion of the sun and keeps the direction of the reflected beam constant.

For the comparison of the dark lines in the solar spectrum with the bright lines of terrestrial sources spectroscopes were furnished with a totally reflecting prism placed in front of the slit and covering half of it. One of the sources was viewed directly and the light from the other was reflected in from the side. The two spectra were seen in the field above one another and thus comparison was easy. It should be noted, however, that with this arrangement the same lines in the two spectra do not coincide exactly in position, unless the beam of light from each source fills the full aperture of the collimator.

The four principal ways of producing spectra are by means of the bunsen flame, the electric spark, the vacuum tube, and the electric arc.

Flame Spectra. The bunsen flame is the simplest method of producing spectra, but it is suitable only for the salts of sodium, lithium, thallium, potassium, barium, strontium, calcium, rubidium, and caesium, and except in the case of the first three of these elements there is difficulty in making the measurements. The usual method of introducing the salt into the flame is on a platinum wire; a small loop is made on the end of the wire, the latter is dipped into hydrochloric acid, and in order to clean it, heated in the flame as long as it gives off a yellow colour; it is then dipped in the salt. The temperature of the bunsen flame 2·5 cms. above the top of the burner varies from 1600° C. in the centre to 1800° C. at the surface. Consequently the platinum wire should be placed in the flame near the surface and tangential to the latter so that the last few millimetres are heated to a bright red.

If the salt will not stay on the platinum loop long enough, it can be placed in a platinum scoop, or a piece of charcoal or bit of asbestos can be impregnated with it, and held in the flame. However, even with these methods the salt often burns away before the observer can get time to measure the lines, and special pieces of apparatus have been devised to give a constant supply of salt. Eder and Valenta employed a wheel on the rim of which was fixed a circle of platinum gauze. Part of this circle dips into a vessel containing a strong solution of the salt and another part projects into the flame. The wheel is rotated by clockwork or a motor and thus constantly carries salt into the flame. Mitscherlich employed little glass tubes closed at the upper ends, into the lower ends of which a wick of asbestos threads or fine platinum wires was fitted, and which were bent so that the wicks projected into the flame. The tubes were filled with the liquid under examination, a little hydrochloric acid or ammonium acetate being added to prevent the formation of crusts on the wires. Gouy's method consists in making the air or gas take up the salt solution in the form of a fine spray, before it reaches the burner. Hemsalech's method consists in making the air go through a chamber in which an electric spark is passing before entering the draught hole at the bottom of the bunsen burner. When the spark passes, the electrodes are disintegrated and their material is carried off by the air current into the flame in the form of a very fine powder.

Much brighter spectra are obtained if instead of the bunsen burner the oxyhydrogen flame is employed.

Sodium is always present in lithium salts as an impurity, consequently when a lithium salt is introduced into the flame the sodium yellow is superimposed on the lithium red. The lithium salts volatilise at a lower temperature than the sodium salts; thus the lithium red extends further into the colder regions of the flame, and the latter has an orange core but a red edge.

According to Kirchhoff and Bunsen the spectroscope can detect $\frac{1}{14000000}$ milligram of sodium in the bunsen flame. Thus spectrum analysis is very much more sensitive than chemical analysis. This extreme sensitiveness to sodium and consequent universal prevalence of sodium in all flames made it more difficult for experimenters to discover the real cause of the yellow lines.

The Electric Spark. If two metal rods are connected to the ends of the secondary of an induction coil and their points are brought near one another, when the coil is started, a stream of sparks passes between them. This is a well-known experiment, and it is used to describe the power of the induction coil; we say, for example, that it gives a six-inch spark.

The induction coil consists of a bundle of straight iron wires termed the core, round which is wound a coil of insulated wire called the primary. The secondary is wound round the primary and consists of a very large number of turns of fine wire insulated with silk and shellac. In the primary circuit there is an interrupter. This in its simplest form consists of a spring to the upper end of which a piece of soft iron is attached on one side; on the other side is a platinum point which rests against a platinum surface. The current passes up the spring and through the platinum point to the surface and then round the primary. This magnetises the core, the core attracts the piece of soft iron, contact between the platinum point and surface is broken, and the current is interrupted. The core then demagnetises, the spring flies back, contact is made and the same cycle takes place over again. Whenever the primary current is made or broken an induced current flows in the secondary. The break takes place much more sharply than the make, consequently the induced electromotive force is much greater one way, and the stream of sparks passes between the points only in the one direction. When the sparks pass in air they have a violet appearance and they give principally the air spectrum.

If, however, one or two Leyden jars or condensers are connected in parallel with the spark gap, the appearance of the spark is changed. Then when the primary is broken, the current in the secondary flows into the condenser, C (fig. 222), until its potential is raised to such a height that the dielectric resistance of the spark gap, G, is broken down; an oscillatory discharge of the condenser then takes place through the gap. In this case the spark gap has to be made much smaller and the spark gives principally the spectrum of the metal points between which it passes.

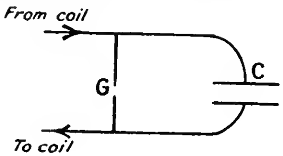


FIG. 222.

The Leyden jar can be charged by a Wimshurst influence machine instead of an induction coil, and then when discharged across a spark gap gives a spectrum in the same way.

The simplest way of obtaining the spark spectrum of a solution is to put it in a test tube, through the bottom of which a platinum wire has been sealed (fig. 223). The platinum wire must not reach the surface of the liquid inside the tube; outside the tube it is connected to the negative terminal of the secondary of the induction coil. The mouth of the test tube is closed with a cork through which passes a glass rod with a platinum wire fused down its centre; the lower end of this wire is a millimetre or so above the surface of the liquid and the upper end is connected to the secondary of the induction coil. The sparks pass between the wire and the surface of the liquid.

The disadvantage of this method is that the sparks attack the glass, and so Hartley has used instead two chisel-shaped wedges of graphite moistened with the solution.


FIG. 223 (from Baly's "Spectroscopy").

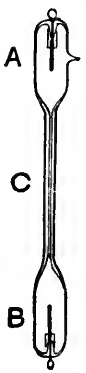


FIG. 224 (from Watson's "Practical Physics").

The Vacuum Tube. Vacuum tubes or Geissler tubes are glass tubes with electrodes sealed through their walls containing gases at a pressure of 1 or 2 mm. or thereabouts. When the electrodes are connected to the secondary of an induction coil the tube lights up.

Figs. 224 and 225 represent vacuum tubes; fig. 224 is the usual form consisting of two wider portions connected by a capillary. Owing to the current being concentrated in the capillary, the tube is much brighter there. The electrodes are made of aluminium wires fused on to platinum wires; the parts going through the glass are platinum, the parts from which the discharge takes place are aluminium. If the electrodes are made solely of platinum a deposit takes place on the glass walls of the tube. Fig. 225 represents the end-on form; the collimator is pointed at it in the direction of the tube C, the capillary is thus viewed end-on and the spectrum is consequently much brighter.

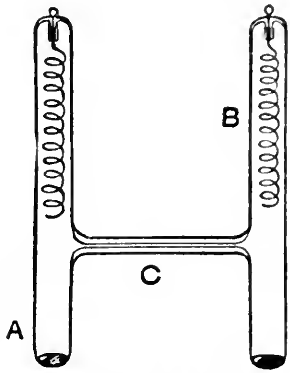


Fig. 225 (from Watson's "Practical Physics").

Unless special precautions are taken, vacuum tube spectra contain impurities due to moisture, dirty glass surfaces, and hydrogen occluded by the electrodes; also the gas itself is gradually decomposed or absorbed by the glass and electrodes.

Formerly the gas in the capillary was thought to have a very high temperature, but the temperature is now known to be very low. On the assumption that all the electric energy reappeared as heat, E. Warburg calculated that in a certain hydrogen tube the temperature could at no point exceed 133° C., while R. W. Wood found by direct measurement with a bolometer that the temperature in a tube of special construction was of the order of 30° C. to 40° C. Wood's tube was formed from the Torricellian vacuum of a barometer, and the bolometer entered through the mercury column.

The Electric Arc. If two carbon rods are connected to the terminals of a battery with an e.m.f. of 80 volts or thereabouts, and if the ends of these rods are pushed together and then drawn apart, a bright discharge passes between them. This is the electric arc. As the current passes, the end of the positive carbon becomes hollowed out into a crater-like depression and the end of the negative terminal becomes pointed. Both the carbons are consumed by the arc, the positive one twice as fast as the negative one. Much the greater portion of the light comes from the positive electrode, less from the negative, and very little from the vapour between the electrodes. The temperature of the positive electrode has been shown by optical methods to be about 4000° C.

If an image of the arc is thrown on the slit of the spectroscope by a lens, by adjusting the latter the spectra of the different parts can be examined in succession. It is found that the spectrum of the electrodes is continuous, while the vapour gives a mass of fine lines. If the lower carbon is the positive one and small quantities of different metals or salts are placed in the crater, they are volatilised and the vapour gives their spectrum. This procedure is, however, not very satisfactory, as the introduction of the salt makes the arc flicker and jump, and besides the salt is soon burned up. It is better to have a hole bored along the centre of the positive carbon and pack it with the salt. Such cored carbons are used in street lighting in the flame arc lamps and can be bought packed with salts of calcium.

In the case of metals such as iron or aluminium the arc may be passed directly between rods of the metal itself, but care must be taken not to have the current too strong, otherwise the positive pole melts. A. H. Pfund * has given the following useful instructions for using the iron arc: The lower electrode must be the positive one and should be about 12 mm. diameter. The upper electrode should be about 6 mm. diameter and project about 3 mm. from a brass bushing which it carries to prevent it becoming too hot. It is best to limit the length of the arc to 6 mm. The arc burns best with about 3·5 amp. on a 220 volt direct circuit. After it is started a bead of iron oxide forms in a cup-shaped depression on the positive electrode, bulging out considerably above the edges of its receptacle. The presence of this bead is essential to the steady working of the arc.

I have found that with plenty of ballast resistance and an induction such as the coil of an electromagnet in the circuit, the iron arc give little trouble. Before starting it the layer of iron oxide which forms on the negative electrode must always be scraped off.

The arc spectrum of mercury can be produced very simply by boring the lower carbon and attaching its lower end by a rubber tube to a mercury reservoir. Then we have only to raise the reservoir sufficiently high to bring the mercury in the core up to the level of the flame. It is cleaner, however, and gives sharper lines, besides preventing the formation of poisonous vapour, to have the arc in a closed tube. Arons was the first to use a closed mercury arc successfully; fig. 226 represents his arrangement. The electrodes are the mercury columns A and B; besides these columns the tube contains only mercury vapour. The arc is struck by inclining the lamp and letting a drop run from A to B. At the electrodes further mercury reservoirs are attached outside to prevent the temperature rising too high.

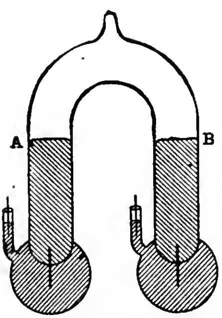


Fig. 226.

The most successful of all mercury arcs has been the Cooper-Hewitt tube, which is used extensively for workshop lighting. The tube takes 3·5 amps. and requires a ballast resistance and an induction. The potential difference between the ends of the tube itself is about 34 volts, and together with the resistance and induction it requires from 50 to 65 volts. The cathode is a little pool of mercury and is at the blackened end of the tube; the anode is an iron ring. Fig. 227 represents the tube together with a mount suitable for laboratory use. Ordinarily the end A rests on the table; to start the arc the mount is rocked over on to the end B and this causes a thread of mercury to

* "Astroph. Jr.," **27**, 1908, p. 296.

flow down the tube and make contact with the iron ring. The disadvantage of the Cooper-Hewitt tube for laboratory use is its size. It

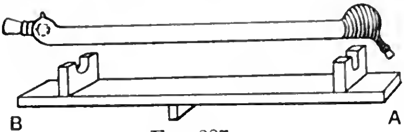


FIG. 227.

is 27 inches long and gives 500 candlepower. But against that it is made in large numbers and is consequently cheaper and more reliable than if it were used solely for spectroscopic work.

The Cooper-Hewitt and Arons arcs are also made in quartz. When working with a quartz mercury arc or with an iron arc the eyes must be protected by glass spectacles, because these sources give off ultra-violet rays which have a harmful effect on the eyes and cause a very painful inflammation. The harmful rays are absorbed by glass.

In addition to the mercury arc there are also mercury vacuum tubes with aluminium electrodes. These contain a few drops of mercury and must be heated to vaporise the mercury, otherwise they do not light up. Vacuum tubes may have also mercury electrodes; the distinction between such tubes and the mercury arc lies then solely in the potential difference and current employed.

Work of Kirchhoff and Bunsen. In 1860 Kirchhoff and Bunsen explained the Fraunhofer lines and asserted that, when a substance was caused to emit light, the spectrum was the sum of the spectra of its components. They thus gathered up and put in a definite form ideas which were, so to speak, in the air at that time, developed them by a series of researches, called attention to their importance, and, in a word, founded spectrum analysis.

The statement that every atom has its own spectrum consisting of lines or bands characteristic of it and produced by no other atom was of immense importance, because, if the spectrum of an unknown substance was examined and the positions of its lines mapped, it was necessary only to compare them with the lines of the known elements, and we could at once obtain the composition of the unknown substance. Moreover, the quantity of substance required for producing the spectrum was very much less than what would be required for chemical analysis. Also the method could be applied to the sun and stars, bodies to which chemical analysis could be applied by no possibility whatever. If a substance under investigation contained lines which could not be identified with the lines of any known element, then the substance contained an unknown element. Bunsen and Kirchhoff at once discovered the elements caesium and rubidium in this way, and were able afterwards to isolate them chemically, and about the same time Crookes discovered thallium by means of its green line.

Balfour Stewart had shown that a body absorbed best the heat rays that it emitted, and the idea of resonance was used by Kelvin in connection with the sodium lines. A vibrating system absorbs the radiations which it emits; thus a tuning-fork is set into vibration if it is placed on the sounding-box of another fork of the same pitch, which is already sounding, and the strings of a piano resound to the sound waves of another musical instrument in its neighbourhood. The strings gain energy at the expense of the waves, which are consequently weakened in intensity. So, if white light passes through sodium vapour, those waves with the same periods as the sodium molecule are absorbed, the other waves pass through with their intensities undiminished, and if the light then enters the slit of a spectroscope, the spectrum is crossed by two dark bands in the yellow. These dark bands have exactly the same positions as the bright lines emitted by sodium in the bunsen flame.

Kirchhoff generalised these ideas, expressed them quantitatively in the statement known as his law, namely, that the ratio of the radiating power to the absorbing power of all bodies is the same for the same wave-length and the same temperature, and applied them with decisive success to the explanation of the Fraunhofer lines. According to his view, which is now universally adopted, the sun consists of a solid or liquid core at a white heat surrounded by an atmosphere at a somewhat lower temperature. The core gives out white light which would show a continuous spectrum, were it not for the fact, that the radiations corresponding to the periods of the molecules in the atmosphere are abstracted in passing through the atmosphere. When a line is absorbed in this way it is said to be reversed and the atmosphere is referred to as the reversing layer. The Fraunhofer lines are thus caused by absorption in the sun's atmosphere. By comparing their wave-lengths with the wave-lengths of the emission spectra produced in the laboratory, the chemical constitution of the sun's atmosphere can be determined. It is found in this way that more than 36 of our elements exist in the sun. Indeed one element, helium, was discovered in the sun before it was found on the earth. Close beside the sodium lines D_1 and D_2 in the solar spectrum there is a line D_3 of wave-length 5875·6 A.U. which formerly could not be produced in the laboratory and which was ascribed by Lockyer and Frankland to a hypothetical element, called helium by them. This element was isolated by Ramsay from cleveite in 1895.

All the Fraunhofer lines are not due to absorption in the solar atmosphere. Some, known as telluric lines, are due to absorption by the water vapour and oxygen of the earth's atmosphere. That some of the lines are due to water vapour was shown by Jansen; he had a large fire lit on one side of the lake of Geneva, examined its spectrum from the other side, and found that absorption lines were produced by the damp air over the lake.

The action of the reversing layer can be illustrated in the laboratory by an experiment that is somewhat difficult to carry out. An arc lamp, such as is used for projecting slides, is taken and the projecting lens is adjusted to form an image of the crater on the slit of the spectroscope. An observer looking into the telescope will then see an intensely bright continuous spectrum. A bunsen burner is then placed to give a sodium flame in front of the slit, and, if the sodium vapour is made sufficiently dense, a dark line can be made to appear in the spectrum. If a screen is placed between the arc lamp and the sodium flame, only the light of the latter enters the instrument and the dark line appears bright. It is not any brighter than before, only in the latter case it is seen against a dark background, whereas in the former case it is seen against a very bright background.

In 1862 Mitscherlich found that it was possible for a compound to give a spectrum; for example, when barium chloride was introduced into the bunsen flame, it gave its own spectrum, not the spectrum of barium and the spectrum of chlorine. Little is known about the spectra of compounds on account of the experimental difficulty of producing them; the compounds are almost always dissociated by the bunsen flame or other means used to produce the spectrum, so that the existence of compound spectra does not really limit the power of spectrum analysis to reveal the elements present in an unknown salt.

Types of Spectra. Spectra are divided into three classes, continuous spectra, band spectra, and line spectra. Solid bodies give continuous spectra; thus the spectra of a red-hot poker, the crater of the electric arc and the filament of an incandescent lamp contain rays of every possible wave-length whatever. This is usually explained by saying that the molecules are so closely packed, that they cannot vibrate in their proper periods. Just when one is beginning to emit regular vibrations another knocks against it, and the vibrations become all irregular again. The oxides of the rare earths, didymium oxide for example, form an exception to the general behaviour of solids; when they are heated they give a line spectrum superimposed on a continuous spectrum.

Band spectra consist of bands separated by dark spaces. When observed with a high resolving power each band is found to consist of very fine lines. The lines become closer and closer on one side of the band until they coincide; this side has consequently a sharp and bright edge called the head of the band. With a low resolving power the bands are not resolved and the spectrum appears channelled or fluted. All spectra of compounds are band spectra.

Line spectra consist of separate bright lines on a dark background or a faint continuous background. The different lines differ in intensity, and when examined with a high resolving power they do not always appear the same. Some are sharp, some are sharp on one side and nebulous or diffuse on the other, and others are diffuse on both sides.

Plücker and Hittorf found in 1865 that the same substance can give two entirely different spectra. This had not been expected by Kirchhoff and Bunsen. The discovery was first made with nitrogen in a vacuum tube, and the results were afterwards extended to a large number of other substances. They all gave two spectra, a band spectrum and a line spectrum. The passage from the band spectrum to the line spectrum took place abruptly when the conditions of the discharge were altered. The band spectrum and line spectrum could not exist together and they had no line in common. Plücker and Hittorf refer to the band spectrum as the first order spectrum and to the line spectrum as the second order spectrum.

Lockyer's Long and Short Lines. If a spark is passing horizontally between two electrodes and an image of it is thrown on to the vertical slit of a spectroscope by means of a lens, then all the lines in the spectrum have not the same length. The slit picks out a vertical section from the spark; some of the lines are emitted only in the centre of the spark where the discharge is most intense and the pressure greatest, while other lines are also emitted far out from the centre where the discharge is weak. The former lines appear short and the latter appear long in the spectrum. This method of studying the effect of the varying circumstances of the discharge is due to Lockyer and is known as the method of the long and short lines.

Ångström's Normal Solar Spectrum. In the early days of spectrum analysis the positions of lines were given on arbitrary scales. Kirchhoff and Bunsen used an arbitrary scale. A. J. Ångström measured the wave-lengths of a very large number of the Fraunhofer lines by means of three diffraction gratings, and in 1868 published a map of the visible spectrum giving the wave-lengths of about 1000 lines. The wave-lengths were expressed in 10^{-8} cms. and carried to two decimal places, whence the name Ångström unit. Ångström had measured the spaces of his grating in terms of a standard metre in the calibration of which an error had been made, and consequently his values were all slightly too small. This error was corrected and the values recalculated by Thalén. Ångström's diffraction gratings were of course plane ones. They were placed normal to the direction of the collimator.

Ångström's normal map standardised spectroscopy and made it possible to determine the wave-length of any line simply by comparison with the solar spectrum and interpolation. Ångström's values remained for many years unexcelled for accuracy.

EXAMPLES.

(1) In calibrating a spectroscope a student finds that sodium bicarbonate gives a line at 53° 4′, potassium chlorate lines at 52° 12′ and 56° 8′, lithium sulphate a line at 52° 36′, thallium sulphate a line at 53° 39′ and a hydrogen vacuum tube lines at 52° 40′, 53° 11′, 54° 19′, 54° 40′, and 55° 14′. Of course some of these lines may be due to impurities. Draw curves representing the

scale readings (*a*) as a function of λ, (*b*) as a function of $1/\lambda$, and (*c*) as a function of $1/\lambda^2$.

(2) Calculate the constants of the Hartmann calibration formula for the above spectroscope, and show by a graph how near the observed values lie to the calculated curve.

(3) In a constant deviation spectroscope the collimator and telescope remain fixed while the prism is rotated and its position read by means of a micrometer screw. The screw is calibrated by means of hydrogen, helium, and mercury vacuum tubes, the readings for the hydrogen lines being 2537, 2187, 1955, for the helium lines, 2548, 2438, and for the mercury lines, 2358, 2208, and 1958. The wave-lengths of the principal lines of helium are 4471·5, 5875·6, and 6678·1 A.U. Draw the calibration curve and calculate the constants of Hartmann's formula for it. Represent Hartmann's formula by a curve and show how closely it fits the observed values.

(4) Show by plotting the dispersion as defined on p. 238 as a function of the wave-length, that it is not constant in the case of a transmission diffraction grating mounted in a fixed position on the table of a spectrometer, although it does not vary so rapidly with the wave-length as in the case of either flint or crown glass prisms.

(5) A spectroscope has a slit which opens only on the one side. A student calibrates it with the lines of Na, Li, Tl, and H, using a constant slit width and placing the cross-wires on the middle of the slit. He then looks for the K violet line, and after widening the slit and expending much time gets a reading on the middle of this line also. Immediately after making this reading he realises that the cross-wires ought always to have been placed on the image of the fixed edge of the slit. What should he do to save his observations?

(6) A spectroscope is being used as a monochromatic illuminator. For this purpose the eyepiece is removed and a cardboard disc with a slit in it inserted in the focal plane of the telescope object glass. The object glasses of the telescope and collimator have the same focal length. Show that for a given intensity of illumination the light is most monochromatic when the collimator and telescope slits are equally wide.

(7) A symmetrical direct vision spectroscope is to be made of the flint and crown glasses for which the indices of refraction are given on p. 61. The prism is to be symmetrical and is to consist of three parts, a flint component in the middle with a crown component on each side. The refracting angle of the flint component is to be such that, if there is an air film between the components, the rays can just enter the flint glass. The D lines are to be undeviated. Calculate the angles of the prisms and the difference in deviation of the two ends of the visible spectrum.

(8) A prism and plane mirror are mounted on the table of a spectrometer in such a way, that the plane bisecting the refracting angle of the prism and the plane of the mirror intersect in the axis of rotation of the table. The combination is known as the Wadsworth mirror prism combination. The rays from the collimator pass through the prism, are reflected by the mirror and then enter the telescope. Show that the rays which pass through the prism at minimum deviation suffer a constant deviation in passing through the combination, and hence show how any spectrometer, which is furnished with a means of reading the position of the prism table, can be converted into a constant deviation instrument of the fixed arm type.

CHAPTER XV.

THE ULTRA-VIOLET.

When silver chloride is exposed to light it darkens in colour, first assuming a violet tint and then becoming dark brown or black. The exact chemical nature of the change occurring is not known, but it has been attributed to partial reduction to metallic silver. In 1801 J. W. Ritter found that this property of light did not stop at the violet end of the spectrum but was greatest beyond the end of the visible spectrum. He thus discovered the ultra-violet spectrum.

Ten years later by projecting Newton's rings in ultra-violet light on silver chloride, Thomas Young showed that the ultra-violet rays were subject to the laws of interference. The diameter of the rings was smaller than the diameter of the rings produced by visible light, and thus their wave-length was smaller. Herschel introduced the name actinic rays for the rays that produced chemical change.

At first the actinic rays were regarded as something radiated from the sun additional to ordinary light and distinct from it. Only gradually the conviction grew that it was the same radiation that produced both the sensation of light and the chemical action on silver chloride.

If an acidulated solution of quinine sulphate is placed in a dark room and a beam of white light is allowed to fall on it, blue light is emitted from the surface of the solution at the point where the beam is incident, also blue light is emitted from the path of the beam inside the solution. The blue light is brightest at the surface decreasing with the thickness of solution through which the beam has passed. If, however, after passing through such a solution the beam falls upon a second solution, it no longer possesses the property of exciting blue light. At first this phenomenon was not properly understood and the blue light was thought to be reflected incident light, but in 1852 there appeared a description of an investigation by Stokes which cleared up the whole matter.

The effect is shown by a great number of other substances besides sulphate of quinine, for example by an alcoholic solution of chlorophyll, paraffin oil, an aqueous solution of fluorescein, uranium glass and fluorspar, and on account of its being shown by fluorspar Stokes introduced the name fluorescence for it. Uranium glass is coloured with oxide of uranium, and flower vases can be bought made of it. They are yellow by transmitted light and fluoresce green, but the glass made for experimental purposes by Schott & Co., Jena, gives a much brighter fluorescence. When a spectrum is projected on a fluorescent screen or on the surface of a fluorescent solution, for example, sulphate of quinine, the blue fluorescent light is emitted by the parts of the surface on which the blue, violet, and ultra-violet portions of the spectrum fall. Thus a fluorescent substance has the property of making an ultra-violet spectrum visible; Stokes investigated the ultra-violet solar spectrum in this way and mapped the positions of the principal Fraunhofer lines in it.

Sulphate of quinine and fluorspar fluoresce with a blue light, chlorophyll with a red light, and uranium glass and fluorescein with a green light. Thus the emitted light is characteristic of the substance and not of the exciting light; the energy of the exciting light is first absorbed by the molecules and then emitted by them as light of another wave-length. Stokes believed that the wave-length of the exciting light must be always less than the wave-length of the fluorescent light, and this fact was known for long as Stokes's law. Modern investigation has shown that it is certainly not always true; Nichols and Merritt have proved, for example, that fluorescence can be produced in the case of fluorescein when the wave-length of the exciting light is greater than that of the centre of the fluorescent band.

Use is made of fluorescence in a well-known experiment used for demonstrating the laws of the reflection and refraction of light to large audiences. A trough with parallel sides of glass is filled with water and a grain or two of fluorescein added. A beam of parallel light from an arc lamp is incident on the surface of the water at B (fig. 228) where it gives rise to a reflected beam BC and a refracted beam BD. At the bottom of the trough at D there is placed a piece of mirror glass which produces the reflected beam DE and consequently the refracted beam EF. If the air were perfectly pure, the paths of the beams AB, BC, and EF would not be visible to an observer at the side; hence dust, for example chalk from the blackboard duster, is scattered in the air, the chalk particles are illuminated by the beams and render the paths visible. The extremely dilute solution of fluorescein in the trough lets through visible light with its intensity undiminished but absorbs ultra-violet light. The energy of this ultra-violet light is re-emitted again as green fluorescent light and consequently the path of the beam in the solution appears a bright green. Every molecule of fluorescein in the path is a source of green light. Care must be taken to get the strength of the solution right, for if there is too much fluorescein in it, the ultra-violet light will be all absorbed before the beam reaches D and only part of the beam BD and none of the beam DE will be visible.

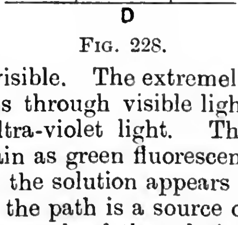


FIG. 228.

Of course dust particles in the water are illuminated by the beam and render its path visible, but they only scatter the incident light and do not shine with a light of their own.

The reason why, when light goes through two solutions of sulphate of quinine in succession, only the first solution exhibits fluorescence is because all the ultra-violet light of the beam is absorbed in the first solution and there is none left for the second.

Stokes used fluorescence as a means of testing the transparency of different substances to ultra-violet light and also for estimating the amount of ultra-violet light emitted by different sources. He found that glass absorbed the ultra-violet light from an alcohol flame, while quartz did not, and that the electric spark was very strong in ultra-violet light.

He had a spectroscope built with quartz lenses and prism, and received the spectra on a uranium glass screen or uranium phosphate screen, and in 1862 published a description of some results obtained with this apparatus. He found that the spectra of the electric arc and spark extended much further than the solar spectrum. In the same year W. A. Miller photographed spark spectra with a quartz apparatus, and found that quartz, fluorspar, and water were all very transparent in the ultra-violet. He also discovered the fact, that for a certain region in the ultra-violet thin films of silver are transparent. From 1874 to 1880 Cornu extended Ångström's normal solar map to the ultra-violet by means of photography with a plane reflection grating. He found that the solar spectrum stopped short at about 3000 A.U; the earth's atmosphere absorbed the radiations beyond this limit. The limit of transmission varied slightly with the season of the year and the height above sea-level.

§ Very many substances fluoresce slightly, but the fluorescence cannot be seen on account of scattered incident light of greater intensity being superimposed on it. To detect it in such cases Stokes used a very ingenious arrangement consisting of a box with two coloured windows, one of which transmitted only the violet and blue while the other transmitted the yellow and red. The substance under investigation was placed inside the box and an intense beam of light allowed to fall on it through the blue window. The blue light scattered or reflected by the object was not transmitted through the yellow window, as the two taken together were quite opaque; consequently the object was not visible through the yellow window unless it gave rise to a yellow or red fluorescence. A very faint fluorescence could be detected in this way.

Fluorescent light is polarised by refraction on leaving the fluorescing body.

Recently R. W. Wood has investigated the fluorescence of sodium vapour and found that, when resolved spectroscopically, it consists of a very great number of fine bands if the fluorescence is stimulated by white light. If, however, monochromatic light of great purity is allowed to fall on the vapour only some of these bands appear, but those that do appear are distributed throughout the whole spectrum. If the wave-length of the exciting light is slightly altered, the bands originally present disappear and others appear. Each group of bands in the complete fluorescent spectrum responds to its own exciting wave-length. Wood terms the spectra produced in this way resonance spectra; they are caused by the sodium molecule resonating to the exciting radiation. The resonance spectra violate Stokes's law in a very marked manner.

§ We have then two means of investigating the ultra-violet spectrum, namely, fluorescence and photography. The best way of employing fluorescence is by means of Soret's fluorescent eyepiece. This consists of a thin plate of uranium glass or other transparent fluorescent substance fixed in the telescope in the position usually occupied by the cross-wires. If an ultra-violet line is focussed on the surface of the plate, it produces a fluorescent image, and this image is viewed through the plate with the ordinary eyepiece of the telescope, which is inclined obliquely to the axis of the telescope, because the image is then seen better. The fluorescent plate is, of course, normal to the axis of the telescope.

A fluorescent eyepiece can be used only with bright spectra and consequently the use of photography is now universal. In Ritter's experiments referred to on p. 256 the silver chloride was exposed to the rays until it actually became black. In 1839 Niepce and Daguerre made known their process, which was a very great advance on what had previously been accomplished. A surface of silver iodide on silver was exposed to the action of light and removed before it presented any visible change; it was then placed over the vapour of slightly heated mercury, and mercury deposited on the parts where the light had acted, more being deposited where the light had been intense and less where it had been weak. Thus if an image had been focussed on the surface, it was represented in every gradation of light and shade. The image had been latent in the surface although not visible until brought out by the mercury.

The modern dry plate consists of an emulsion of silver bromide in gelatine, which has been poured while warm on to a large glass plate and allowed to set. The glass plate is then cut into smaller sizes. The light produces no visible action on the plate until the developer is poured over it. Then the particles of silver bromide on which the light was incident are reduced to black metallic silver. After development the plate is "fixed" by immersion in a solution of hyposulphite of soda, which dissolves away all the sensitive particles on which the light has not acted. It is then washed and dried. As the bright parts of the image, those on which most light falls, come out black on the plate and the dark parts come out light, the picture is said to be a negative. By placing it in contact with sensitive paper and allowing light to act through the negative on the paper, an image called a positive is produced on the paper. In this image the light and shade are correctly rendered; it is the picture wanted, the negative being only an intermediate step.

In photographing spectra we always stop at the negative because it is a matter of indifference as to whether the lines come out white on a black background or black on a white background. Also it is an advantage to have the picture on glass, because it is often necessary to measure the distance between the lines and glass does not stretch the way paper does. The same reason applies against making the negative on films. In films the gelatine containing the silver bromide is spread on a roll of celluloid instead of glass.

17 *

The slower plates produce a finer grained image than the more rapid ones. The ordinary dry plate is sensitive to light from 5000 A.U. to 2200 A.U. with a maximum of sensitiveness in the violet. Very strong lines may come out below 2200 A.U., but the gelatine of the plate begins to absorb at 2500 A.U. and at 2200 A.U. the light does not penetrate more than ·02 mm. into the gelatine film, i.e. to a greater distance than about one-tenth of its thickness.

Vogel found in 1873 that if the plate is bathed in a solution of one of certain dyes it becomes sensitive to some of the rays absorbed by this dye. Use has been made of this discovery to make plates sensitive to the green, yellow, and red. The ordinary commercial orthochromatic or isochromatic plates, which are made by putting some eosin or erythrosin into the emulsion, are sensitive as far as the yellow. The well-known "Wratten Panchromatic Plate," made by Messrs. Wratten & Wainwright, is prepared by bathing the finished plate in a solution of certain of the isocyanines, and while as sensitive to the ultra-violet as the ordinary non-colour sensitive plate, it is sensitive to the whole visible spectrum to beyond the red lithium line. It does not give the red potassium line.

The development of ordinary plates is carried on in a dark room illuminated solely by a red lamp. In the case of orthochromatic plates care must be taken with regard to the spectral purity of the red light. In the case of the panchromatic plate the red light affects the plate and consequently cannot be used. The plate must be developed in absolute darkness, or, if this is not possible, a faint green light may be used. Green is chosen because the eye is most sensitive to green and the plate is nearly equally sensitive to all colours, hence by using green we obtain a better illumination for the same damage to the plate.

§ As was discovered by Stokes glass absorbs the far ultra-violet light. The limiting radiation transmitted varies with the kind of glass and the distance traversed by the light in it; some flint glass prisms absorb almost to the end of the visible spectrum. Crown glass lets through more; a plate 2 mm. thick will transmit perhaps to 3200 A.U. Recently Schott & Co. of Jena have put a new optical glass named "Uviol" on the market, which when in a thickness of 1 cm. passes light of wave-length 2970 A.U. With a quartz prism and lenses the last line transmitted is the strong aluminium one at 1852 A.U. This line is best detected with a fluorescent ocular since, as is mentioned above, ordinary gelatine plates do not go beyond 2200 A.U. Fluorite has been found by Schumann to transmit to 1000 A.U., while Iceland spar passes light up to 2150 A.U.

Fluorite of optical quality is extremely rare, and the high birefringence of Iceland spar and the brittleness of its surfaces make it unsuitable, so that quartz is the material almost universally used for work in the ultra-violet. If a prism is made of a doubly refracting material such as quartz or Iceland spar, each incident ray is resolved into two when it enters the prism, the ordinary ray and the extraordinary ray, consequently in general there are two spectra produced, the ordinary and extraordinary spectra. If, however, the prism is an isosceles one and the optic axis is parallel to its base, the ordinary and extraordinary rays coincide for the wave-length that passes through the prism at minimum deviation and at this wave-length the ordinary and extraordinary spectra coincide accurately with one another. If the rays make a small angle with the optic axis in going through the prism, then the images they form do not quite superimpose. The double refraction of quartz is so small that if the middle of the spectrum is set at minimum deviation, the whole spectrum from 7000 A.U. to 2200 A.U. can be obtained sharp on one plate; the want of coincidence at the ends of the two spectra is not sufficient to cause appreciable error. It is, however, otherwise for a substance with a high double refraction like Iceland spar.

When a beam of light traverses a quartz prism in the direction of the optic axis, it decomposes into a right-handed and a left-handed circularly polarised beam, and these have slightly different velocities. It is this difference of velocity which causes quartz to rotate the plane of polarisation of a plane polarised beam traversing it in the direction of its axis. Now difference of velocity means difference of refractive index. Thus if an isosceles prism is made solely of right-handed quartz with the axis parallel to the base, it is impossible to set it at minimum deviation for any wave-length for the two circularly polarised components simultaneously. This error due to the optical rotation of the quartz is quite distinct from and exists in addition to the error due to the double refraction which was discussed in the last paragraph. As shown by the calculation on p. 223 it is a small error, and it can be eliminated by using a prism of a special type known as the Cornu prism. This latter consists of two 30° prisms, one of right-handed quartz and the other of left-handed quartz, cut with the optic axes perpendicular to the faces in contact, and placed together so as to form a 60° prism. The difference of deviation between the two beams produced on entering the first prism is removed on emerging from the second prism.

In a quartz spectrograph, as an instrument used for photographing the spectrum is called, the lenses are generally single quartz lenses. They could be made achromatic by combining them with fluorite lenses, but it is not worth the cost Now the index of refraction of quartz varies from 1·614 at 2313 A.U. to 1·539 at 7685 A.U., and consequently the focal length of an uncorrected quartz lens varies 13 per cent over the same range. In a quartz spectrograph the collimator is adjusted so as to make the rays in the middle of the spectrum go through the prism parallel. The extreme ultra-violet rays are then convergent and the red rays divergent. Hence the distances of the foci of the extreme ultra-violet and the red from the object glass of the telescope differ by much more than 13 per cent, and to get the whole spectrum sharp at once the surface of the photographic plate must be inclined at an angle of about 20° to the axis of the telescope. The collimator and telescope lenses are cut with their axes parallel to the optic axis of the quartz, one out of right-handed quartz and the other out of left-handed quartz.

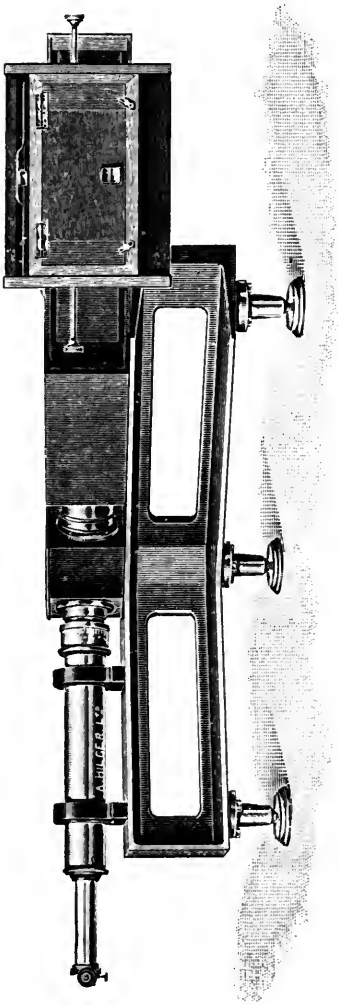


Fig. 229.

Most quartz spectrographs are made by A. Hilger, Ltd., London, and fig. 229 shows one of their best-known types. The collimator is to the left and is of the usual type, the prism is closed in by a wooden box and the telescope is replaced by a wooden camera. The slide is very clearly shown; it can be racked up vertically by a screw at the side so that several spectra may be taken above one another on the same plate.

It is, of course, possible to fit a small camera on to a telescope in place of the eyepiece and by this means take photographs of small regions of the spectrum at a time. In this case, since the regions are small, the photographic plate may be set at right angles to the axis of the telescope.

The Féry Spectrograph. This spectrograph, the invention of Prof. C. Féry, is the outcome of an attempt to do away with the lenses altogether and perform their function by giving the faces of the prism suitable curvatures.

Let NMPQ be the prism and let A be the centre of curvature of the face PQ. Then AP and AQ are normals to this face. Draw a circle AC through A touching the face of the prism PQ at its mid point. The points P and Q are on the face of the prism, but, since the linear dimensions of the surface of the prism are small in comparison with the diameter of the circle, they may be considered also to be on the circle. The radius of curvature of the face PQ is equal to the diameter of the circle.

Let C be a source of light and CP, CQ two rays to the prism. The angles of incidence of these rays are equal because they stand on the same arc AC. Consequently the angles of refraction are also equal, and the refracted rays PM and QN when produced backwards must intersect at a point B on the circle ACQ. The position of B varies with

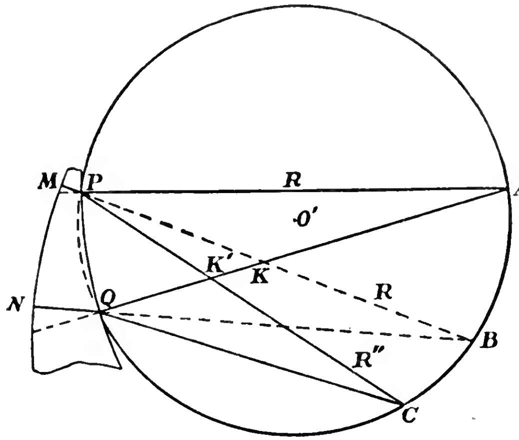


Fig. 230.

the colour of the light. Let B be the centre of curvature of MN, the back surface of the prism. This surface is silvered by the tin-foil mercury process. It reflects the refracted rays back on their path, and consequently on leaving the prism they converge again to C for one particular colour. If the prism is made of quartz and the optical axis is approximately parallel to NQ and MP, the doubling produced by the optical rotation on entering the prism is removed on emerging from it. Of course, since the above construction holds for any two rays provided that the linear dimensions of the prism are small in comparison with

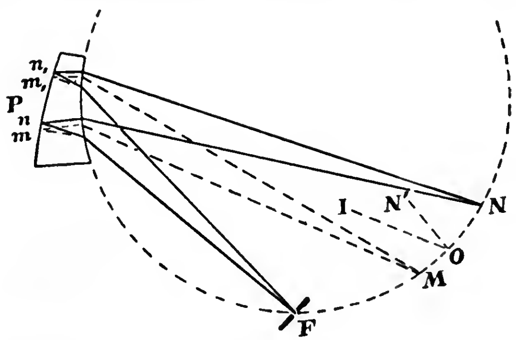


Fig. 231.

the radii of curvature of its surfaces, it holds for the whole pencil of rays incident on the prism from C.

Now consider fig. 231. F represents a slit. The refracted rays are no longer incident normally on the silvered surface of the prism but make a small angle with the latter, consequently they are not reflected back on their own path and the image of the slit is not formed on the slit at F but on the circle to the side of F, its distance from F varying with the colour. The red rays are shown by the full lines and come to a focus at N, while the extreme ultra-violet rays are shown by the dotted lines and come to a focus at M. Thus a sharp spectrum is formed along the arc MN without the use of lenses.

The angle between the faces MN and PQ is roughly 30°. The angle ION′, which the incident light makes with the normal to the photographic plate, is about 51°.

The Féry spectrograph in its action has some resemblance to the Rowland concave grating. The degree of approximation of the circle to the surface is the same in both cases. Like the Rowland grating it works at a smaller aperture than the quartz spectrograph does, at $f/30$ instead of $f/19$, and hence gives fainter spectra. (In the ordinary notation of photography $f/30$ means that the diameter of the effective aperture of the lens is one-thirtieth of its focal length.)

A Simple Means of Photographing Spectra. Any student possessing an ordinary snapshot or stand camera can with a little patience and ingenuity take quite good photographs of spectra with it. All that is required in addition to the camera is a Thorp grating replica with about 14,000 lines to the inch costing 15 or 20 shillings. The grating is mounted immediately in front of the camera lens. No collimator is necessary. The source of light, an electric spark for example, is placed at a distance of say 10 feet, and all the rays from it are parallel enough without the intervention of a collimator lens. Or instead of a spark the capillary of a vacuum tube may be taken, the thicker portions of the tube being screened off. The capillary acts then as a slit. The only disadvantage of an apparatus like this is the shortness of the spectra. If in the formula $\lambda = e \sin \theta$ we substitute in succession $7\ 10^{-5}$ and $3{\cdot}3\ 10^{-5}$ cms. for λ, which would correspond to the limits reached with a Wratten panchromatic plate through a glass lens, we obtain 22° 48′ and 10° 31′ for θ. The focal length of a quarter plate camera lens is about 5 inches; hence the length of the first order spectrum in the case of a quarter plate camera would be roughly $\frac{10 \times 5}{57}$, or about $\frac{7}{8}$ of an inch. Of course a prism is less suitable than a grating owing to the dispersion of its spectrum being less than the dispersion of the ordinary grating spectrum. With the arrangement described here the exposures are seldom more than 30 seconds.

Absorption Spectra. If light from an incandescent gas mantle is focussed on the slit of a spectroscope and a plate of blue cobalt glass placed in the path of the rays, the continuous spectrum of the mantle is seen to be crossed by three dark bands, a broad one with its centre at about 5300 A.U. and two sharper ones with their centres about 5900 A.U. and 6500 A.U. These bands are said to form the absorption spectrum of cobalt glass. Their position in the spectrum depends on the composition of the glass, and their width on the percentage of cobalt oxide present and the thickness of the plate. Generally speaking, the regions between the bands are not very transparent, though the red side of the band in the red is. The absorption spectrum of a solution can be obtained in the same way by putting it in a test tube and holding it in front of the slit. When examined in this way a dilute aqueous solution of cobalt chloride shows a broad band in the green, and a dilute aqueous solution of potassium permanganate shows five dark bands in the green. The absorption bands of solutions and solids are, as a rule, very broad and ill-defined, occupying large regions of the spectrum. Aqueous solutions of salts of didymium and erbium and the other rare earths, however, show comparatively sharp bands, as do also their crystals; the latter at the temperature of liquid air show absorption lines comparable with the emission lines of gases in sharpness. Glass coloured with didymium oxide has a very interesting absorption spectrum consisting of sharp bands in the yellow and green with transparent regions between. Gases and vapours show absorption line and band spectra comparable with their emission spectra in fineness.

In mapping the absorption spectra of solutions it is usual to put

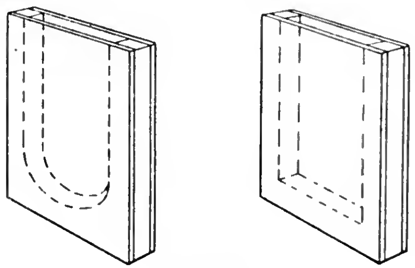


FIG. 232.

them in cells with parallel sides such as are shown in fig. 232. The best background for use in the visible spectrum is the incandescent mantle. Its image should be thrown on the slit with a lens, but slightly out of focus, so as not to show the detail of the mantle in the spectrum. If a metal filament lamp is used as source, its light must be focussed on a ground glass plate, which is placed in front of the slit with just sufficient space between it and the slit for the cell with the solution. Without the ground glass in this case the continuous background will not appear the same brightness the whole way up.

There is difficulty in mapping the position of an ill-defined broad absorption band. One way of proceeding consists in moving the cross-wires into the band from one side until they can no longer be seen against the dark background and reading the position, then repeating the operation from the other side, and finally taking the mean of the two readings. It is found, however, when the breadth of the band is less, either owing to the solution being weaker or to the length of the path in it being shorter, that the readings taken in this way do not always give the same result. They do so only when the absorption increases at the same rate on both sides of the band.

Hence it is usual to employ a method introduced and used by Hartley, which depends on photography and is valid for the whole range of the spectrum to which photography is applicable. This method is best explained by the consideration of a special case, and for this purpose one of Hartley's diagrams is reproduced on p. 267. The curve represents the absorption of nitric acid. A standard solution was prepared by dissolving a certain quantity of the acid in water, and photographs were taken of a spectrum through layers of this solution 5, 4, 3, 2, and 1 mm. thick. The photographs were next examined and observations made of the positions of the edges of the regions which transmitted no light. These positions were then entered up in the diagram. They are represented by the abscissæ and are measured in oscillation frequencies—Hartley used oscillation frequencies instead of wave-lengths—an oscillation frequency being the number of wave-lengths in the millimetre. Thus the numbers at the ends of the scale, 2850 and 4200, correspond respectively to wave-lengths of $3{\cdot}51\ 10^{-5}$ cms. and $2{\cdot}38\ 10^{-5}$ cms., and the region represented lies wholly in the ultra-violet. The ordinates represent thickness of solution and the hollow in the curve represents an absorption band. Thus, when the solution was 5 mm. thick, there was an absorption band extending from 3087 to 3830 with a very narrow transparent interval at the latter point. When the solution was 4 mm. thick the band extended from 3087 to 3674 and the transparent region from the second point to 3896. When the solution was 3 mm. thick the band extended from 3087 to 3647 and the transparent region to 3919. When the solution was less than 1 mm. thick the band ceased to be visible. To find the edges of the absorption band and transparent region for any thickness it is only necessary to erect an ordinate equal to it, and draw a horizontal line through its upper end. Its intersections with the curve are the required points. Thus the curve gives a satisfactory representation of the absorption.

When the absorption of solutions of thallium nitrate and silver nitrate of equivalent strength to the nitric acid is represented in the same way, it is seen that they have absorption bands at the same place in the spectrum but very much blunter. A solution of potassium nitrate of the same strength gives exactly the same curve as nitric acid. Hence the absorption band must be due to the NO_3 radical, but its intensity is influenced by the atomic weight of the base to which the latter is attached.

The spectrum of the incandescent gas mantle does not go far into the ultra-violet, consequently as a background for absorption spectra in the ultra-violet it is necessary to use another source. Hartley employed the electric spark between electrodes made of an alloy containing

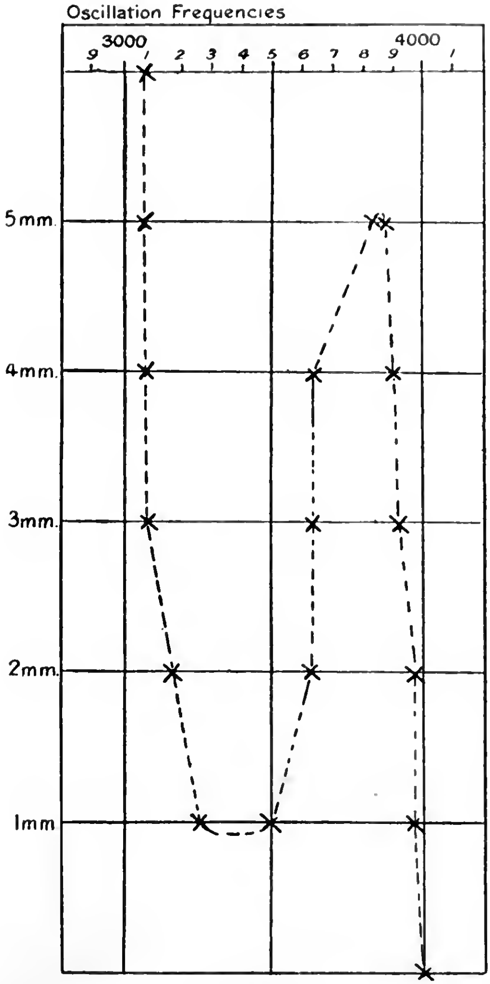


FIG. 233.

lead, tin, cadmium and bismuth. The spectrum of this alloy contains a large number of sharp lines with a faint continuous background. The wave-length or oscillation frequency can be determined by recognising the lines. E. C. C. Baly, in applying the same method, has used the iron arc as background. It has more lines and is much brighter, but owing to the complexity of its spectrum it is less easy to determine the wave-length in it.

Hartley's method has been applied in much detail to the absorption spectra of organic compounds. It is found that these spectra are additive but with a strong constitutive influence. They have been of the greatest service in many cases in elucidating chemical constitution. But, in spite of all the work done, it has hitherto been impossible to form a clear physical picture of the connection between absorption bands and chemical constitution.

Schumann's Work in the Extreme Ultra-Violet. It has already been stated that the gelatine of the photographic plate absorbs strongly at 2200 A.U. and that quartz does not transmit below 1850 A.U. Schumann, who first explored the region below 1850 and who used fluorite as the material of his prism and lenses, found that the air of the atmosphere absorbed the rays and that it was necessary to build a vacuum spectrograph from which the air was excluded. A layer of air 1 mm. thick at 76 cms. pressure absorbs entirely all the rays below 1700 A.U. By means of photographic plates containing silver bromide with just a trace of gelatine Schumann found that of all the substances examined the spectrum of hydrogen extended furthest, reaching to about 1000 A.U.

EXAMPLES.

(1) A quartz spectrograph has one 60° Cornu prism and the object glasses of collimator and camera are single lenses, the focal length of each being 30 cms. for Na light. The prism is set permanently so that light of wave-length 3404 A.U. is transmitted at minimum deviation. Make an accurate drawing to scale of the curve along which the images of the diffeernt lines come to a focus.

(2) Map the absorption spectra of solutions of potassium permanganate and cobalt chloride and determine the wave-lengths of the maxima of absorption. Also determine the position of the absorption bands in cobalt glass. (All blue glasses are coloured with cobalt oxide; the position and intensity of the bands depend on the constitution of the glass to which the oxide is added.)

CHAPTER XVI.

SPECTROSCOPY. LATER WORK.

By using a new principle Rowland succeeded in making a screw the pitch of which was accurate to 10^{-5} inch. This was much in advance of what had previously been accomplished. With this screw he built a dividing engine for ruling gratings and ruled gratings with as many as 43,000 lines to the inch, but found that better results were obtained with a smaller number, 14,438 lines to the inch. His first paper on the subject was published in 1882. The ruling point was a diamond. The glass gratings made were not so good as the metal ones and they wore down the ruling point more, so Rowland devoted his attention to reflection gratings ruled on metal surfaces.

Besides greatly increasing the accuracy of the dividing engine Rowland, as has been mentioned in Chapter X, made the great advance of ruling the grating on a concave metal surface instead of on a plane one. Such a grating gives the spectrum in focus along the arc of a circle without the use of lenses. It thus eliminates all trouble due to absorption of light in the lenses; it also eliminates chromatic error and the consequent labour involved in focussing. Also, as concave gratings were made with radii of curvature as great at 21 feet, they enabled spectra to be taken on a scale quite unprecedented.

It has been shown (p. 178) that the resolving power of a grating is Nn where N is the total number of rulings and n the order of spectrum observed in. The resolving power at any point in the spectrum is defined as $\lambda/d\lambda$, where λ and $\lambda + d\lambda$ are the wave-lengths of two lines which can just be seen apart. Rowland succeeded in ruling 110,000 lines on a breadth of $5\frac{1}{2}$ inches. In the second order this gives a resolving power of 220,000. In order to achieve the same result with a flint glass prism in the yellow part of the spectrum it would be necessary for the base of the prism to have a length of about 220 cms., which of course is not practicable.

The process of ruling the Rowland gratings required very much patience and skill. It took months to make a perfect screw for the ruling engine and longer to find a suitable diamond point. The dividing engine was kept in the basement of the Physical Laboratory of the Johns Hopkins University in Baltimore. It was driven by a water motor. The plate was placed on the engine, then the experimenter left the room, waited until the temperature which had been disturbed by his entrance recovered its normal value, and then started the engine from outside. The room was not entered until the work was finished. When all went well it took five days and nights to rule a 6-inch grating having 20,000 lines to the inch.

Owing to the high cost of ruled gratings Thorp contact copies or celluloid grating replicas, as was mentioned on p. 174, are now used universally for ordinary laboratory work. Mr. Thorp has even found it possible to mount his replicas on a spherical glass surface, ruled side next the glass, and thus make a concave grating. When mounted on glass such a grating is not suitable for the far ultra-violet, as the rays are absorbed in passing through the glass. Rayleigh has successfully copied ruled gratings by photography.

With the concave grating as with other gratings the spectra of the different orders are superimposed. Thus, for example, in the case of a plane transmission grating on which the light is incident normally, the wave-length corresponding to the deviation θ is given by $n\lambda = c \sin \theta$, and the D_1 line, which has a wave-length of 5896 A.U., coincides in the first order with a line in the second order which has a wave-length of 2948 A.U. With the plane grating the two lines are not in focus at the same time owing to the chromatic error of the lenses, but with the concave grating both lines are, and coincidences between lines of different orders can be determined with great accuracy. In this way Rowland determined the wave-lengths relatively to the D_1 line of lines in various parts of the spectrum. He then photographed the solar spectrum and by means of these lines he was able to attach a scale to the photographs giving the wave-lengths of the intermediate lines.

All Rowland's wave-lengths depend thus on the D_1 line. Its wave-length was determined by Bell with two glass and two metal Rowland gratings by measuring the deviation and evaluating the grating space in terms of the standard of length. His result was 5896·18 A.U. in air. Rowland corrected this value, and combining it with the results of previous determinations by other observers adopted 5896·156 A.U. in air for his final value. He published a list of the wave-lengths of 1100 lines for all of which he considered the error to be less than ·01 A.U. Rowland's values superseded Ångström's and until recently were taken as the basis of wave-length measurement.

It is mentioned on p. 150 that by means of his interferometer Michelson evaluated the metre in terms of the red, green, and blue radiations of cadmium. His results give of course at the same time the wave-lengths of these lines. By means of their interferometer Fabry and Perot found the ratio of the wave-lengths of a large number of lines in the visible spectrum to Michelson's standard radiations and thus determined the wave-lengths of the former. One of the lines measured in this way was the D_1 line. Their value for it was 5895·932 A.U. in air, which differs widely from Rowland's value. Their values for the other lines also differ from Rowland's, and the ratio between their result and Rowland's result for the same line is not constant.

This discrepancy aroused much discussion, and in 1904 Kayser attacked Rowland's method of coincidences. Two of Rowland's largest gratings were used and the ratio between the wave-lengths of two lines determined experimentally with each grating by means of the method of coincidences. If the one wave-length was assumed, the results for the otner differed by ·03 A.U., which is distinctly greater than would have been expected. The diffraction grating thus does not seem suitable for absolute work. Its use as a means of determining wave-length must in the future be confined to interpolation between lines the wave-lengths of which have been determined by interferometer methods. But as regards resolving power and convenience in working it holds the same place in our estimation as before.

§ A stretched string gives out different notes at the same time, the fundamental and its overtones, and the same is true of an organ pipe or a vibrating plate. In these cases mathematical relations are known to exist between the fundamental and the overtones. For example, in the case of the stretched string their frequencies are in the ratio of the natural numbers. Now, whatever view we adopt as to the origin of spectra, there is no doubt they are due to a vibrating system in some way characteristic of the atom or molecule. It seemed natural to expect that the different lines in a spectrum might be caused by the different modes of vibration of the same system, and thus arose a search for a mathematical relation between their wave-lengths.

At first investigators were misled by the acoustical analogue and sought to show that the frequencies were to one another as simple whole numbers. Schuster proved that these early results were merely chance coincidences, and progress was not made until the introduction of Rowland's gratings and the great increase in the accuracy of wave-length determination which they rendered possible. The first to achieve a decisive result was Balmer. In 1885 he showed that the wave-lengths of the first nine lines of the hydrogen spectrum can be represented by the formula

$$\lambda = \mathsf{A}\frac{n^2}{n^2 - 4},$$

where A is a constant and n an integer which varies from 3 to 11. The formula could not be verified further as there were no further measurements available then. Afterwards additional lines were discovered in the vacuum tube spectrum of hydrogen by Ames and also in the spectrum of solar protuberances and in the so-called flash spectrum by Evershed. At a solar eclipse, when the sun is obscured with the exception of the edge where the reversing layer is, the spectrum of the latter, i.e. the Fraunhofer lines, flashes out bright for two or three seconds and is known as the flash spectrum. These additional lines are represented by the formula with remarkable accuracy. This is shown by the following table which is taken from a paper by Evershed.

n	Observed.	Calculated.	Difference.
3	—	6563·07	—
4	4860·9	4861·52	− 0·6
5	4341·0	4340·63	+ 0·4
6	4102·3	4101·90	+ 0·4
7	—	3970·22	—
8	3889·21	3889·20	+ 0·01
9	3835·60	3835·53	+ 0·07
10	3798·05	3798·04	+ 0·01
11	3770·78	3770·77	+ 0·01
12	3750·25	3750·30	− 0·05
13	3734·54	3734·51	+ 0·03
14	3722·04	3722·08	− 0·04
15	3712·14	3712·11	+ 0·03
16	3703·99	3704·00	− 0·01
17	3697·22	3697·29	− 0·07
18	3691·71	3691·70	+ 0·01
19	3687·05	3686·97	+ 0·08
20	3682·93	3682·95	− 0·02
21	3679·48	3679·49	− 0·01
22	3676·43	3676·50	− 0·07
23	3673·84	3673·90	− 0·06
24	3671·53	3671·48	+ 0·05
25	3669·52	3669·60	− 0·08
26	3667·70	3667·82	− 0·12
27	3666·15	3666·24	− 0·09
28	3664·71	3664·82	− 0·11
29	3663·40	3663·54	− 0·14
30	3662·14	3662·40	− 0·26
31	3661·16	3661·35	− 0·19
∞	Theoretical Limit	3646·13	

The first column gives the value of n, the second the observed value of λ, the third the value of λ calculated by the formula, and the last column the difference between the observed and calculated values of λ. As n increases the lines become closer and closer together and also fainter, finally disappearing in a continuous band which probably represents the remainder of the series. Theoretically there should be an infinite number of terms in the series and the last term, which is obtained by making n infinity, should be at 3646·13 A.U. The constant* of the formula was calculated from the first three terms and is given by

$$\frac{1}{A} = 27418{\cdot}75.$$

§ Two or three years after the publication of Balmer's work Rydberg and Kayser and Runge started a systematic search for similar series in the spectra of other elements. Rydberg used the data of other workers; Kayser and Runge, who worked together, mapped the spectra afresh by means of a Rowland grating with great care and

* The formula gives the wave-length in vacuo; to get the third column above the result has been divided by the index of refraction of air.

thoroughness. Also in order to reduce their results to vacuum they measured the index of refraction of air throughout the whole visible and ultra-violet spectrum. Rydberg found that there were series in the spectra of the other elements somewhat similar to the hydrogen series and that these series could be represented by the formula

$$\lambda^{-1} = A + \frac{B}{(n + \mu)^2},$$

where A, B, and μ are constants and n runs through all positive integral values starting either with 1 or 2. Kayser and Runge discovered the existence of the same series independently and simultaneously but represented them by the formula

$$\lambda^{-1} = A + Bn^{-2} + Cn^{-4},$$

where A, B, and C are constants and n runs through all positive integral values starting at 3.

Balmer's formula can be written in the form

$$\lambda^{-1} = \frac{n^2 - 4}{An^2} = \frac{1}{A} - \frac{4}{An^2}$$

and hence is a special case of Kayser and Runge's formula. If the $(n + \mu)^2$ in Rydberg's formula is taken to the numerator and expanded as a series, that formula becomes

$$\lambda^{-1} = A + Bn^{-2} - 2B\mu n^{-3} + \ . \ . \ .$$

and so it also includes Balmer's formula as a special case. The essential difference between Rydberg's formula and that of Kayser and Runge is the substitution of the third power of n for the fourth. Both formulæ are approximate; they do not represent the positions of the lines quite within the error of observation. Kayser and Runge's formula is on the whole the more accurate, though Rydberg's has found more favour with subsequent workers. Kayser and Runge regard their formula as the first terms of the expansion in a series of an unknown function. Rydberg claimed that the B in his formula has the same value for all elements, but this is only approximately true and the B in Kayser and Runge's formula also does not vary much.

The following table gives the constants of Kayser and Runge's formula for the different elements as determined by Kayser and Runge and by Runge and Paschen :—

	Nebulous Series or First Subordinate Series.			Sharp Series or Second Subordinate Series.		
	A.	B.	C.	A.	B.	C.
Li	28587	109629	1847	28667	122391	231700
Na	24492	110585	177	24549	120726	197891
K	21991	114450	111146	22022	119393	62506
Rb	20939	121193	134616	20899	113556	76590
Cs	19743	122869	305824			
Cu	31592	131150	1085060	31592	124809	440582
Ag	30712	130621	1093823	30696	123788	384303
Mg	39796	130398	1432090	39837	125471	518781
Ca	33919	123547	961696	34041	120398	346097
Sr	31030	122328	837473	31066	118044	296136
Zn	42945	131641	1236125	42955	126910	632850
Cd	40755	128635	1289619	40797	128635	555137
Hg	40160	127484	1252695	40218	126361	613268
Al	48308	156662	2505331	48244	127527	687819
In	44515	139308	1311032	44535	126766	643584
Tl	41543	132293	1265223	41506	122617	790693
O	23208	110388	4814	23194	107567	63108
S	20087	109598	113556	20078	108745	18268
Se	19266	108901	94293	19287	111960	1227

	Principal Series.		
	A.	B.	C.
Li	43584	133669	1100084
Na	41543	130233	800791
K	35091	127207	623087
Rb	33762	125521	562255
Cs	31509	125395	486773

If we consider first the alkalis Li, Na, K, Rb, and Cs, we find that each of them has three series in its spectrum, two subordinate series and a principal series. But each of these three series itself is double. For each value of n there are two lines, the less refrangible of which is given by the above constants. In the principal series the two lines corresponding to the same value of n get closer together as n increases, and superimpose on one another when n is large. The D lines of sodium are the members of the principal series for which $n = 3$. The red line of lithium is the member of the principal series for which

$n = 3$. In the case of lithium the two lines corresponding to the same value of n are always coincident.

In each of the subordinate series the two lines with the same value of n have always the same difference in their frequencies, and this difference is the same for both subordinate series. Each subordinate series can thus be regarded as two series for which B and C are the same but A different. The two values of A for the sharp series are approximately the same as the two values of A for the nebulous series. The brighter lines are contained in the principal series. The lines of the second subordinate series are usually sharper but fainter than those of the first. In the spectra of the alkalis almost all the lines are contained in the series. This is not the case with the other elements.

Cu and Ag behave like the alkalis, but only the first terms of their principal series have been found.

Mg, Ca, Sr, Zn, Cd, Hg have all six series, each six falling into two groups of three. The three series of each group have the same values of B and C but different values of A. The table gives the value of A for the least refrangible series. The corresponding series of each group have approximately the same value of A.

Al, In, Tl have each four series, i.e. two double series which behave like the subordinate series of the alkalis. The green line of thallium belongs to the second subordinate series and is the less refrangible line for which $n = 3$.

O, S, and Se seem to have a first subordinate series of triplets and a first subordinate series of pairs and also a second subordinate series of triplets and a second subordinate series of pairs. The constants are for the least refrangible member of the triplets.

If the wave-lengths be calculated from the table, it will be found that in each group of the periodic system the lines shift towards the red with increasing atomic weight, but from group to group as a whole the lines shift towards the violet.

There have been many attempts to explain the origin of spectral series but they are all regarded as unsuccessful.

Doppler's Principle. The position of a line in a spectrum can be altered by relative motion of the source and observer in the line of sight. The possibility of this was pointed out by Doppler in 1842. Doppler's views on the subject were neither accurate nor clear and met with much opposition. He thought that the velocity of stars with a continuous spectrum could be determined from their colour, that those approaching rapidly should look blue and those receding rapidly should look red, while the others should appear white. This is not the case, because in a continuous spectrum, if some of the radiations move into the ultra-violet, others come out of the infra-red into the visible spectrum and so no change of colour occurs. The effect on the position of spectral lines of relative motion of observer and source in the line of sight was first treated properly by Fizeau, and the principle is thus often called the Doppler-Fizeau principle, especially in France.

If the observer is at rest and a source with the period τ is moving towards him with velocity v, then in one period the source approaches him by a distance $v\tau$. If V is the velocity of light in the medium in question, the ordinary value of the wave-length is $V\tau$, but if the source is moving towards the observer each wave obtains a start on the preceding one of $v\tau$ and the wave-length is consequently reduced to $V\tau - v\tau$ or $(V - v)\tau$. It is thus given by

$$\frac{(V - v)}{V} \lambda,$$

where λ is the value for no relative motion in the line of sight, and the lines in the spectrum of the source are consequently displaced towards the violet. If the source is moving away from the observer the sign of v is changed and the lines are displaced towards the red. The values of V and τ are, of course, not changed by the motion.

If the source is at rest and the observer moving towards it with velocity v it can easily be seen by a diagram that he will meet v/λ more wave-lengths in a second. If he were standing still he would receive V/λ wave-lengths per second. Consequently he receives $(V + v)/\lambda$ wave-lengths instead of V/λ, the frequency is apparently increased in the ratio V to $V + v$, the wave-length has the apparent value

$$\frac{V}{V + v} \lambda,$$

and the spectral lines are displaced towards the violet. If the observer is moving away from the source the sign of v alters and the lines are displaced the other way.

The effect of motion of the source on the apparent frequency of a note is easily observed in acoustics. For example, if a railway engine passes an observer at the side of the line sounding its whistle, the observer hears the pitch of the whistle fall as the engine passes. Suppose that the true pitch of the whistle is 500 and the speed of the engine 60 miles an hour or 88 feet per second. The velocity of sound is 1100 feet per second. The apparent pitch of the whistle when the engine is approaching is consequently

$$\frac{1100}{1100 - 88} 500 = \frac{1100 \times 500}{1012} = 543,$$

and the apparent pitch of the whistle when the engine is receding is $\frac{1100}{1188} 500 = 463$. When the engine is passing, therefore, the apparent pitch of its whistle suffers a very marked change. If we suppose that the whistle is stationary and the observer passes it at 60 miles per hour, then by the second formula the apparent pitch changes in this case from 540 to 460. The change is approximately the same in both cases.

The velocity of light is so enormously greater than the velocity of sound that a velocity of 60 miles per hour of the source does not alter the wave-length of spectral lines appreciably. The first decisive evidence that Doppler's principle could be applied to spectroscopy was obtained from astronomy in the case of the rotation of the sun. It can be shown from the motion of the sun-spots on its surface, that the equatorial zone of the latter is rotating with uniform angular velocity, its apparent period as seen from the earth being 27·25 days. The period increases with the distance of the zone from the equator. The sun's visible surface does not rotate as a solid. There are currents on it like those of our atmosphere and ocean. The radius of the sun is 433,000 miles, and the linear velocity of a point on its equator is about 1·25 miles per second. If the spectrum of the point on the approaching edge is observed, all the Fraunhofer lines should be displaced towards the violet, and if the spectrum of the point on the receding edge is observed, they should be displaced towards the red. Then if $d\lambda$ is the change in the wave-length produced by the motion

$$\lambda - d\lambda = \frac{V - v}{V}\lambda, \quad \text{or} \quad d\lambda = \frac{v\lambda}{V}.$$

This gives $\dfrac{\lambda}{d\lambda} = \dfrac{186,000}{1{\cdot}25} = 150,000$ approximately, which is well within the resolving power of a large grating. By observations on sun-spots, then, we are able to calculate what the displacement of the Fraunhofer lines should be, and hence by direct measurement of the latter to verify the theory. The displacement of the Fraunhofer lines has been determined by several experimenters, the most accurate measurements being those made by Dunér (Upsala). As a reference mark he took two telluric lines, two dark lines due to absorption in the earth's atmosphere, and compared them with two iron lines close beside them due to the sun's reversing layer, using in succession the spectrum from the two edges of the sun. The difference in the distance between a telluric and a solar line in the two cases was found, as required by the theory, to be that due to a velocity of twice the value for the sun's edge.

The smallest velocity that can be determined by Doppler's principle is about $\frac{1}{4}$ ml. per sec. In the yellow this means a shift of ·008 A.U.

Spectroscopic Binaries. One of the most interesting applications of Doppler's principle has been to the discovery of double stars so close that no telescope can resolve them, but which are proved to be double by the behaviour of the lines in their spectra. Two cases occur. In the first case the lines of the spectrum exhibit a periodical shift about a mean position. This is caused by a bright star and a dark star rotating round one another. When the bright star is approaching the lines move the one way; when it is receding they move the other way. In the second case the lines double and undouble periodically. This is caused by two stars of approximately equal brightness rotating round one another. When the lines appear double, one is approaching and the other receding. When they appear single, the one is in front of the other, and they are both changing the direction of their velocity in the line of sight.

Such stars are known as spectroscopic binaries. Those of the second kind can be detected without using a collimator, by taking the star itself as slit and having the prisms in front of the object glass, as is done by Pickering. He uses an eleven or fourteen-inch object glass with four large prisms in front of it, each large enough to cover the whole lens. The clockwork which makes the telescope follow the motion of the star is made to go a little fast or slow so as to give the spectrum breadth. The refracting edges of the prisms lie east and west. With this arrangement the stars in the field appear as spectra upon the photographic plate.

About one-seventh of all the stars investigated with the spectroscope has been found to be double.

Doppler's principle was applied successfully to the steady motion of single stars several years before the discovery of spectroscopic binaries. In 1867 Sir William Huggins showed the feasibility of the method for the case of Sirius.

§ Comets have usually a faint continuous spectrum on which are superimposed certain bright bands. The continuous spectrum is in part due to reflected sunlight. The bands are the same as those given by the blue cone at the base of a bunsen burner, which are always found when hydrocarbons are burned in air. They are termed the Swan spectrum. Their cause is not certain; they have been ascribed by different workers to carbon dioxide, carbon monoxide, and acetylene.

The spectra of stars have been divided by Secchi into four classes: (1) those which are continuous but on the top of which the hydrogen lines appear very intense and reversed, i.e. black like the Fraunhofer lines; (2) those with a spectrum resembling the sun; (3) those which are continuous but with dark bands superimposed, the bands being sharply defined at the more refrangible edge; and (4) those which are continuous but with dark bands superimposed, the bands being sharply defined at the less refrangible edge.

In 1864 Sir William Huggins found bright lines in the spectra of certain nebulæ, thus proving that they were gaseous and not aggregations of stars. About half the nebulæ show a bright line spectrum, which is in all cases substantially the same and consists of two hydrogen lines and two lines in the bluish-green which belong to an element as yet undiscovered on the earth.

Broadening of Spectral Lines. With an ordinary one prism spectroscope the spectral lines are simply coloured images of the slit. They appear of a uniform brightness from edge to edge, and when the breadth of the slit is halved, their breadth is halved. In the case of a single spectral line all the radiations have not the same wave-length. If we take for example the green line of thallium, the wave-length of which is 5350·7 A.U., it may contain radiations with wave-lengths varying from 5350·5 A.U. to 5350·9 A.U. Thus, even if the slit were infinitely narrow, the line would still have a finite breadth itself of 0·4 A.U. Most single prism spectroscopes can just show the sodium lines double, and their difference of wave-length is 5·96 A.U., so that under ordinary conditions with such instruments the breadth of the line is much smaller than the breadth of the slit.

It is quite otherwise with instruments of high resolving power such as the Rowland grating. In the photographs which they give the breadth of the line itself is greater than the breadth of the slit. Consequently the individual characters of the lines make themselves visible and the lines are all found to be brightest in the middle and to decrease in brightness towards the edges and also to have widely varying breadths.

When the pressure in the source is low the diminution of intensity towards the edge of the line is due to Doppler's principle. According to the kinetic theory of gases the vibrating particles have velocities of translation. The components of these velocities in the line of sight, or the radial velocities as they are called, have all possible values from zero to infinity. The number of particles with a given radial velocity diminishes rapidly with the magnitude of that velocity. Although the vibrations have all exactly the same period, this period suffers an apparent change owing to the radial velocity, and the ray is refracted to a slightly different point in the spectrum, the number of rays with a given deviation diminishing very rapidly with the magnitude of the deviation. According to Rayleigh the brightness of the line should be given by

$$e^{-k\phi^2},$$

where ϕ is the distance from the centre of the line in angular measure and k is a constant. k depends on the mean velocity of translation, diminishing with the temperature and increasing with the mass of the system. The fact that at the same temperature the lines of the first subordinate series are nebulous and those of the second subordinate series are sharp would seem to show that the two series are emitted by particles carried by two systems of different mass.

It has been shown experimentally by Michelson that, when the pressure in the source is higher than $\frac{1}{1000}$ atmosphere, the breadth of the lines is also increased by the impacts between the molecules. Between two impacts the source emits a regular train of sine waves. At the impact there is a sudden change of phase. It will be shown in Chapter XXI that a train of sine waves with a constant period but irregular changes of phase is equivalent to the superposition of a number of perfectly regular trains, the periods of which differ slightly from the period of the irregular train. The broadening due to this cause does not require to be symmetrical.

The Mechanism of the Spark. Mohler photographed the spectrum of an electric spark which passed between electrodes of different metals in the direction in which the instrument was pointing, then reversed the direction of the electrodes and took another spectrum on the same plate. A small displacement was obtained corresponding to an average velocity of the source of about ·37 km./sec. Previous to this Schuster and Hemsalech had photographed the spectrum of a spark on a film which moved with a velocity of 100 metres per second in a direction perpendicular to the spectrum lines. The spark passed parallel to the slit. They found that the air lines were perfectly straight showing that the spark first passed through the air with great rapidity. The metal lines were curved showing that the metal vapour started from the electrodes and moved with diminishing velocity towards the centre. The velocity of travel of the vapour varied from 1·3 to ·4 km./sec. Photographs taken with the prism removed showed that the discharge was an oscillatory one, and that the first spark which was a very rapid one passed solely through the air. It vaporised the metal and the metal vapour conducted the other sparks.

Stark's Work on Canal Rays. Recently (1905) Stark has made an important application of Doppler's principle to vacuum tube spectra.

If in a highly exhausted vacuum tube a perforated kathode is used, rays emerge from the holes in the kathode in the direction away from the anode. These rays are called canal rays. They are deflected by a magnetic and by an electric field. From the deflection it can be shown that they consist of particles travelling with a high velocity, that their mass is of the order of the hydrogen atom, and that they carry a positive charge.

Stark used a cylindrical glass tube of 4 or 5 cms. diameter. The kathode was an aluminium disc pierced with many holes 1 mm. broad. The gas in the tube was hydrogen. A prism spectrograph was pointed towards the end of the tube so that the canal rays came directly towards it. The canal ray particles have in general not all the same velocity; hence if they emit a spectrum the displacement due to the radial velocity should vary with the magnitude of that velocity, and while the spectral lines should be displaced towards the violet, they should at the same time be widened out. This is what Stark found. The maximum velocity of the canal ray particles can be calculated from the potential difference they pass through in front of the kathode and the result agreed with that obtained from the displacement of the lines.

If λ_n is the wave-length of the light emitted from the particles normally to their line of flight and λ_p the wave-length emitted parallel to their line of flight, then from p. 276,

$$\lambda_p = \frac{V - v}{V}\lambda_n,$$

where v is the velocity of the particles and V the velocity of light. This gives

$$\frac{\lambda_n - \lambda_p}{\lambda_n} = \frac{v}{V}.$$

Stark found that the hydrogen lines H_β, H_γ . . . showed the Doppler effect, and that the above expression was constant for all the lines of the series. They were thus all due to the same positively charged system. The band spectrum of hydrogen showed no Doppler effect and was hence due to an uncharged system.

The velocities of the canal ray particles were very large, the maximum being 600 km./sec.

Reflection of Light from a Moving Mirror. Let us suppose that light of wave-length λ and velocity V is incident at an angle θ on a plane mirror, and that the mirror is moving forward in the direction of its normal with velocity v. The component of the velocity of the mirror in the direction of the ray is $v \cos\theta$, the velocity with which the waves arrive at the mirror is consequently $V + v\cos\theta$, and the number of waves received per second is increased in the ratio of V to $V + v\cos\theta$. In the same way the reflected ray is only leaving the mirror with a velocity of $V - v\cos\theta$, the waves emitted by the mirror in a second are spread over a distance $V - v\cos\theta$ instead of a distance V, and the number of waves per unit length of the ray is increased in the ratio of $V - v\cos\theta$ to V. Combining both effects we find therefore that the effect of the reflection has been to diminish the wave-length in the ratio

$$\frac{V}{V + v\cos\theta} \quad \frac{V - v\cos\theta}{V}.$$

Since v is small compared with V this reduces to

$$\frac{1}{\left(1 + \frac{v\cos\theta}{V}\right)}\left(1 - \frac{v\cos\theta}{V}\right)$$

$$= \left(1 - \frac{v\cos\theta}{V}\right)^2 = 1 - \frac{2v\cos\theta}{V}.$$

It can be shown by Huygens' principle that owing to the motion of the mirror the angle of reflection is not exactly equal to the angle of incidence, but the difference is very small, and its effect on the change in the wave-length can be neglected.

If instead of a mirror we have a rough surface, which diffuses the incident light in all directions, then the wave-length of the light scattered in a direction making an $\angle\phi$ with the normal can obviously be obtained by the above reasoning, if we substitute ϕ for θ in the expression for the velocity with which the ray leaves the mirror.

The above theory has been employed by Wien in deriving what is known as the Wien displacement law in the theory of complete radiation. It has been verified experimentally by Galitzin and Wilip and has been applied in astrophysics to determine the angular velocity of the planets. As Doppler's principle is assumed in the above theory of the moving mirror, the verification of the latter is at the same time a verification of the former.

Experimental Verification of Doppler's Principle. Doppler's principle was first verified experimentally in the laboratory by Bélopolsky. He used multiple reflection from mirrors mounted on the rims of wheels which were revolved at a high speed. The verification was repeated in 1907 by Prince Galitzin and J. Wilip with the same apparatus but with the substitution of an echelon spectroscope in place of Bélopolsky's spectroscope. The echelon spectroscope gave a much greater dispersion. Six reflections were used and the mirrors had a linear velocity of 30 metres per second.

The Rotation of the Planets. The planets are dark bodies illuminated by reflected light from the sun. Their spectra consequently show the Fraunhofer lines. Owing to their rotation the different parts of their surfaces have different radial velocities. If their spectra are observed, there will be a displacement of the lines depending on the part of the surface from which the light comes, and, if the displacement is measured for different parts of the surface, it will be possible to determine the angular velocity of the planet.

This method has been applied amongst other cases to Venus and to Saturn's rings. From observations of spots on its surface the earliest observers assigned to Venus a period of about twenty-three hours, but Schiaparelli considered that his observations make it certain that the rotation was very slow, and that it was probable that Venus like Mercury always kept the same face towards the sun, and had therefore a period of 225 days. The question has been settled by the spectroscope. The earlier spectroscopic observations seemed to point to the short period but they did not stand investigation, and it is now certain that the long period is the correct one.

The planet Saturn is surrounded by three thin flat concentric rings in the plane of its equator as is shown in the diagram. The

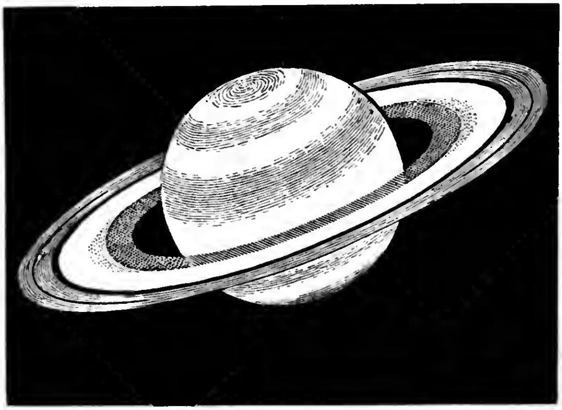


Fig. 234.

question arose as to the constitution of these rings. Two hypotheses were possible, one that they were solid, the other that they consisted of a swarm of small particles, each pursuing its separate course, packed so closely that they appear to be continuous.

Clerk Maxwell showed in 1857 by a mathematical investigation that the first hypothesis was untenable, that thin solid rings would not be dynamically stable, that if they received a very slight displacement they would break up. In 1895 Keeler obtained spectroscopic proof that the inner edge of the rings rotated faster than the outer edge. This decides in favour of the second hypothesis, because if the rings were solid, their angular velocity should be constant, while if they consist of particles each particle would be kept in its orbit by the attraction of the planet itself. The acceleration towards the centre, namely, $\omega^2 r$, should thus be proportional to $1/r^2$ where r is the distance of the particle from the centre of the planet. This makes ω proportional to $1/r^{3/2}$ and consequently greater at the inner edge. The values of the velocities obtained spectroscopically agreed with the values required by the mathematical theory.

The Pressure Shift. In 1895 Humphreys and Mohler discovered at Baltimore that, when a source of light is subjected to a high pressure, the lines in the spectrum are shifted towards the red. Their source of light was an electric arc enclosed in an iron cylinder with a quartz window. The pressure was increased by pumping air into the cylinder and the greatest pressure used was about fifteen atmospheres. The spectrum was photographed with a large concave grating. The shift is directly proportional to the increase of pressure. It is also proportional to the wave-length of the line. Thus all shifts can be reduced for purposes of comparison to a standard pressure and wave-length. When they are reduced in this way to a pressure of twelve atmospheres and a wave-length of 4000 A.U. it is found that they range from 24 to 132 thousandths of an Ångström unit. The value is the same for the lines of the same series.

The shift is independent of the temperature. The lines of those substances which have in the solid state the greatest coefficients of linear expansion have the greatest shifts. The converse is also true. Band spectra are unaffected by increase of pressure.

Of course, when the density of the sodium in a bunsen flame is increased and the partial pressure of the sodium vapour consequently increased, the D lines broaden. The shift detected by Humphreys and Mohler is proportional to the absolute pressure and independent of the partial pressure of the vapour and the breadth of the lines.

Structure of Spectral Lines. When spectral lines are examined with a high resolving power, many of them are seen to be complex. For example, the red hydrogen line is a doublet, the distance between the components of which is ·14 A.U. Other lines have fainter narrower lines, usually referred to as satellites, close beside them. The green line of mercury, 5460·7 A.U., is a line of this type. Owing to the ease with which it can be produced, it has been investigated very often indeed. Fig. 235 is a photograph of it taken by Prof. J. C. McLennan * with an echelon spectroscope.

* Proc. Roy. Soc. A, **87**, p. 276, 1912.

The broad line in the photograph is the main component of the line. To the left of it is a strong satellite which appears double and outside that a fainter one; on the right there is a broad one which

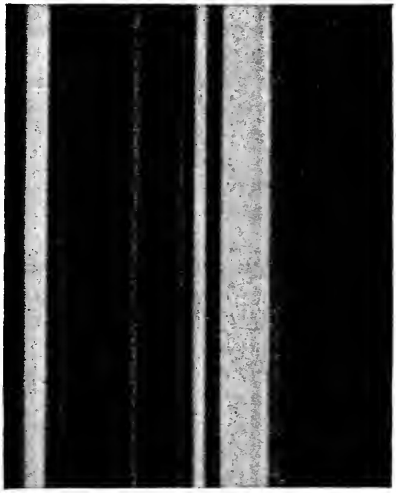


FIG. 235.—The green line of Hg.

runs together with the main line and close beside it a fainter one. There is also another satellite on this side much further out; it is at a distance of ·243 A.U., or less than $\frac{1}{20}$ of the distance between the D lines, from the central component. The other lines in the photograph belong to the spectrum of the next order.

For the structure of a line to be revealed the vapour pressure in the source must be small. In the case of the mercury arc in air, for example, the central component of the 5460·7 line would be broadened so much that all the satellites would be obscured.

The Zeeman Effect. Faraday made an investigation on the possible effect of a strong magnetic field upon a spectral line. A sodium flame was placed between the pole pieces of an electromagnet and the appearance of the D lines examined when the field was on and off. His results were wholly negative.

This experiment was repeated in various ways by successive observers but positive results were not obtained until thirty-four years later, in 1896, by which time the resolving power of spectroscopic apparatus had greatly increased. In that year Zeeman used a large electromagnet, the pole pieces of which were drilled so that observations could be made on the light emitted from the source in the direction of the lines of force, and he found at first that, when the field was on, the spectral lines were broadened. From theoretical considerations he expected that the magnetic field would change the line into two circularly polarised lines of slightly different wave-length and opposite direction of rotation, and that the line which was rotating in the direction of the magnetising current would have the shorter wave-length of the two. To test this he placed a quarter wave plate and nicol in the path of the beam, and arranged them so that no right-handed circularly polarised light could get through. Then when the cross-wire was placed on the line and the direction of the magnetising current reversed, the line shifted. When the field was on and the line was viewed in the direction of the lines of force, he found thus that its two edges were circularly polarised in opposite directions and in the way required by theory.

Later work with increased resolving power showed that the effect of the magnetic field was to change the original line into separate lines. In the simplest case, when viewed in the direction of the magnetic field, instead of the original line two lines were seen, each of which was circularly polarised, the direction of rotation of the more refrangible line being the same as the direction in which the magnetising current encircled the core of the magnet. In this same case when the light emitted by the source at right angles to the magnetic field was examined, the line was found to be divided into three separate lines. The two outside components had the same wave-length as the two lines seen in the direction of the magnetic field. They were equidistant from the middle component and were both plane polarised in a plane parallel to the magnetic field, while the middle component had the same wave-length as the original line and was plane polarised in a plane at right angles to the magnetic field. The line thus became a doublet when viewed along the lines of force and a triplet when viewed at right angles to the lines of force. As the latter is the easier way of observing the effect, it is employed much the more frequently, and this case is consequently referred to as the "normal triplet".

In the majority of cases the resolution is more complex. Some lines when viewed at right angles to the field give, for example, doublets, quartets, and sextets. The separation of the components is always proportional to the field strength. Each member of a spectral series shows the same type of subdivision and, when measured in frequencies, the separation between its various components is always the same for the same field strength.

A successful explanation of the Zeeman effect, as the phenomenon has been called after its discoverer, has been given for the case of the normal triplet.

Let us suppose that we have a particle of mass m carrying a charge e measured in electromagnetic units, and that this particle is vibrating about a point. Take the point as origin. Then, if in the usual way we consider each component of the particle's motion separately, its equations of motion may be written

$$\frac{d^2x}{dt^2} + n^2x = 0, \quad \frac{d^2y}{dt^2} + n^2y = 0, \quad \frac{d^2z}{dt^2} + n^2z = 0 . \qquad (1)$$

We suppose that it is acted on by a force towards the origin proportional to the distance from the origin. The solution of these equations is, as may be found by substitution,

$$x = \mathsf{A} \cos (nt + \alpha),\ y = \mathsf{B} \cos (nt + \beta),\ z = \mathsf{C} \cos (nt + \gamma), \tag{2}$$

where A, B, C, α, β, and γ are constants all independent of one another. The solution thus represents three simple harmonic motions with the same period $2\pi/n$ but different amplitudes and different phases. It can be shown that the particle always moves on the surface of an ellipsoid and keeps to one plane, so that its orbit is an ellipse.

Now suppose that an electric field of intensity H acts on the particle in the direction of the z axis. It is shown in books on electromagnetism that when a current of strength i electromagnetic units is flowing in a thin wire at right angles to a magnetic field of strength H, then on every element of length ds of the wire there is a force of magnitude $\mathsf{H}i\,ds$ at right angles both to the direction of H and the direction of the current. If the left hand be held with the thumb and index finger and middle finger at right angles to one another, then if the index finger gives the direction of H and the middle finger the direction of i, the thumb gives the direction of the force. The rule does not hold for the right hand. A single particle with a charge e* and velocity v is equivalent to a current element ids; hence it experiences a force $\mathsf{H}ev$ at right angles to H and to its line of motion.

In the problem the velocity of the particle has three components, $\frac{dx}{dt}, \frac{dy}{dt}$, and $\frac{dz}{dt}$. Their effects can be considered separately. The z component of the velocity is parallel to H and hence causes no force. The other two components cause forces $\mathsf{H}e\frac{dy}{dt}$ and $-\mathsf{H}e\frac{dx}{dt}$, if the rule of signs be considered, parallel respectively to the x and y axes. The z equation of motion remains unaltered, while the x and y equations become

$$\frac{d^2x}{dt^2} + n^2x = \mathsf{H}\frac{e}{m}\frac{dy}{dt}$$

and

$$\frac{d^2y}{dt^2} + n^2y = -\mathsf{H}\frac{e}{m}\frac{dx}{dt}.$$

Try as solutions $x = \mathsf{D}_1 \cos (n_1t + \phi)$ and $y = -\mathsf{D}_1 \sin (n_1t + \phi)$. Then, when the common factors are cancelled out, both equations reduce to

$$-n_1^2 + n^2 = -\mathsf{H}\frac{e}{m}n_1.$$

In the same way if we try $x = \mathsf{D}_2 \cos (n_2t + \theta)$ and $y = \mathsf{D}_2 \sin (n_2t + \theta)$ both equations reduce to

$$-n_2^2 + n^2 = +\mathsf{H}\frac{e}{m}n_2$$

The first of these equations can be written

* In electromagnetic units.

$$n_1^2 - \mathrm{H}\frac{e}{m}n_1 + \left(\frac{\mathrm{H}e}{2m}\right)^2 = n^2 + \left(\frac{\mathrm{H}e}{2m}\right)^2$$

or
$$n_1 - \frac{\mathrm{H}e}{2m} = \sqrt{n^2 + \left(\frac{\mathrm{H}e}{2m}\right)^2}.$$

Since the change produced in the period by the magnetic field is a very small one, $\left(\frac{\mathrm{H}e}{2m}\right)^2$ can be neglected in comparison with n^2. Hence

$$n_1 = \pm n + \frac{\mathrm{H}e}{2m}.$$

But n_1 must have the same sign as n. We therefore obtain finally

$$n_1 = n + \frac{\mathrm{H}e}{2m}.$$

Similarly
$$n_2 = n - \frac{\mathrm{H}e}{2m}.$$

Our first solution represents uniform motion in a circle of radius D_1 with a period of $2\pi/n_1$, and the second solution uniform motion in the other direction with a period of $2\pi/n_2$ in a circle of radius D_2. The directions are marked in fig. 236.* An electric current flowing in the direction of the n_2 rotation would produce a magnetic field with the same direction as H, and the n_2 rotation has the greater frequency, if we make the assumption that e is negative, that the charge on the particle is a negative one. The complete solution is obtained by adding the two solutions. Hence if we suppose that each of these circular motions gives rise to a circularly polarised wave, then the doubling of the line as viewed along the direction of the lines of force is completely explained.

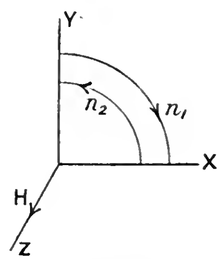


FIG. 236.

Suppose, now, that the observer is viewing the source at right angles to the lines of force, that, for example, he is looking at the particle from X. Light waves are transverse so that he sees only the Y and Z components. They are given by

$$x = -\mathrm{D}_1 \sin(n_1 t + \phi),$$
$$y = \mathrm{D}_2 \sin(n_2 t + \theta),$$
and
$$z = \mathrm{C}\cos(nt + \gamma).$$

He thus sees three lines, the periods of which are given by $2\pi/n_1$, $2\pi/n$, and $2\pi/n_2$. The two Y lines, i.e. the outside lines, are only half the intensity of the middle line, because, when all the vibrating particles are taken into consideration, they were together equal to the Z line in intensity before the field was put on. If we assume, as is usual, that the particle vibrates at right angles to the plane in which the wave is polarised, the Y vibrations are polarised in the direction of the field and the Z vibration at right angles to this direction.

* Z is drawn outwards from the XY plane.

If the velocity of light be denoted by c, the wave-lengths of the three components are given by

$$\lambda_1 = \frac{2\pi c}{n_1},\ \lambda = \frac{2\pi c}{n} \text{ and } \lambda_2 = \frac{2\pi c}{n_2}.$$

Hence
$$\lambda_1 - \lambda = 2\pi c\left(\frac{1}{n_1} - \frac{1}{n}\right) = \frac{2\pi c}{nn_1}(n - n_1)$$
$$= \frac{2\pi c}{n^2}\frac{He}{2m},$$

if the sign of e is neglected, since n_1 is approximately equal to n. Similarly we find that $\lambda - \lambda_2$ is equal to the same expression. If we write $d\lambda = \lambda_1 - \lambda = \lambda - \lambda_2$ we obtain from the last equation

$$d\lambda = \frac{\lambda^2}{2\pi c}\frac{He}{2m}.$$

Thus the distance between the components is proportional to H.

The value of $d\lambda$ is, of course, always very small. The D_1 line becomes a quadruplet in the magnetic field, and Zeeman found that for $H = 10{,}000$ c.g.s. units the distance between the outer components of this quadruplet was only $\frac{1}{13}$th of the distance between the D_1 and D_2 lines. By determining the average value for the normal triplets in the spectrum of mercury at a field of 24,600 c.g.s. units, Runge and Paschen found that e/m has the value $1{\cdot}6\ 10^7$. Other observers have found results in accordance with this. Now in electrolysis the ratio of the charge on the hydrogen atom to the mass of that atom is $\frac{1}{1600}$ times this value. If we make the assumption that the charge is the same in both cases, the mass of the vibrating particle must be $\frac{1}{1600}$ times the mass of the hydrogen atom. The sense of the rotation of the two circularly polarised lines seen in the direction of the lines of force shows that the charge is a negative one.

Negatively charged particles with approximately the same value of e/m have been found in other fields of investigation. The kathode rays are a stream of such particles, the β rays of radium are a similar stream with a greater velocity, which is nearly equal to the velocity of light. The same particles are also emitted from metals under the action of ultra-violet light, and calculation shows that they are the cause of the absorption of light by certain colouring matters. They thus seem to be a unit of which the atoms are built, at least in part, and they have been given the name of electrons.

The Zeeman effect makes it clear that the spectral lines are due to the vibrations of electrons, but unfortunately it takes us only a certain way. It has hitherto proved impossible to form a definite picture as to how the quartets and other complex forms of resolution are caused, and we have been compelled to fall back upon a general theory. According to Lorentz, in these other cases the electron vibrates freely before the magnetic field is imposed, but afterwards it is subject to constraints and has as many "degrees of freedom" as it has components in the magnetic field. When the magnetic field is removed, these degrees of freedom reduce to the ordinary three, for all of which the period is the same. It has not yet been found possible to form any satisfactory picture of the constraints.

In order to produce the Zeeman effect a powerful electromagnet must be employed. The highest field worked with has been about 33,000 c.g.s. units. To obtain this the pole pieces must be pointed and their tips brought close together with just room for the source of light between. As source of light the vacuum tube or electric spark has been used. The spark must pass in the direction of the field itself, otherwise the field deflects it and puts it out. The resolving power of the spectroscope must also be of the highest. Concave gratings have been used, especially in the second order, but for the purpose of the study of the Zeeman effect and the structure of fine lines in recent years a number of new high power spectroscopes has been designed.

The Echelon Grating. The most important of these is the echelon grating spectroscope, which was invented by Michelson and described first in 1898. The resolving power of a diffraction grating is given by Nn where N is the number of rulings and n the order observed in. With the usual concave grating n cannot be greater than 2 or 3, hence to increase the resolving power N must be made as large as possible. But there is a limit to the size of the area that can be covered with uniform rulings. In the echelon grating N is small, seldom more than 30, but n is made exceedingly large.

Fig. 237 represents an echelon grating. It is made of ten plates all equally thick and arranged with each plate projecting the same distance beyond the one which comes after it. It is used with a telescope and collimator, the object glasses of which are large enough to fill the end face of the echelon completely with light. The arrows represent the incident and emergent rays. Let t be the thickness of the plates, s the width of each step, and μ the index of refraction of the glass. If we consider the emergent rays from any two successive steps, the one set has traversed the distance in air which the other set has traversed in glass. Hence they have a path difference of $(\mu - 1)t$. But by Huygens' principle every point on the faces of these steps can be considered as a secondary source. Thus all the steps can be regarded as sending out rays in all possible directions. If we consider the two rays represented by the dotted arrows, their path difference will be greater than $(\mu - 1)t$ and the path difference will increase with the obliquity. If the light is of wave-length λ, whenever the path difference is equal to $n\lambda$, where n is an integer, the rays from the different steps reinforce, and we have a bright line just as in the case of the diffraction grating. But in the diffraction grating the smallest path difference between suc-

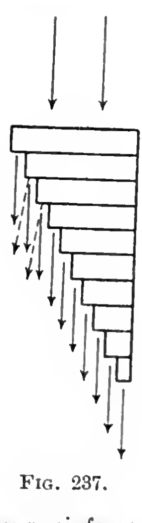


FIG. 237.

cessive rulings was 0. Here it is $(\mu - 1)t$; hence if $\mu = 1{\cdot}52$ and $t = 1$ cm., which is its usual value, and the D_1 radiation is under observation, the smallest value of n is

$$\frac{(\mu - 1)t}{\lambda} = \frac{\cdot 52}{5{\cdot}896 \times 10^{-5}} = 8800.$$

This is consequently the order of the first spectrum observed. A very slight inclination of the diffracted rays is sufficient to increase the path difference by λ; hence the successive orders are very close together. If the echelon has 30 plates, its resolving power in the neighbourhood of the D lines is 264,000. Many echelons have been made with approximately this value; they can separate lines which are apart only $\frac{1}{264}$ of the distance between the sodium lines.

The great advantage of the echelon is, that since the width of each step, i.e. the distance s, is usually about 1 mm., the diffraction maximum for the light going through each step is a very narrow one (cf. p. 179); consequently most of the rays go straight through and all the light is thrown into one or two orders, and these very high orders. The spectra are consequently very bright. The disadvantage is, that the spectra overlap to such an enormous extent that the appearance in the field cannot be interpreted, unless an auxiliary spectroscope is used to purify the light before it falls on the echelon collimator slit. Fig. 238 represents a small echelon suitable for placing on the table of an ordinary spectrometer.

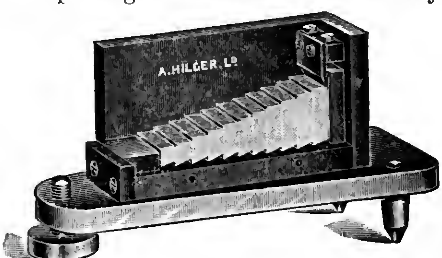


FIG. 238.

All the plates in an echelon must agree in thickness to less than $\frac{1}{20}$ of a wave-length. They are all cut from a single plate which is previously figured to the required accuracy.

Determination of Difference of Wave-length by the Echelon Grating. Let A and B be two corresponding points on successive steps. Consider the rays which make an $\angle\theta$ with the normal. Draw BC perpendicular to AC. The path difference is equal to

$$\mu BD - AC.$$

By projecting the sides of the quadrilateral ACBD on AC we find that

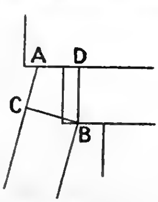


FIG. 239.

$$AC = BD\cos\theta - AD\sin\theta = t\cos\theta - s\sin\theta$$
$$= t - s\theta,$$

since θ is small. Substituting this value for AC in the expression for the path difference and equating the result to $n\lambda$, we obtain

$$n\lambda = \mu t - t + s\theta$$

or

$$n\lambda = (\mu - 1)t + s\theta.$$

Differentiate with respect to λ. This gives

$$n = t\frac{d\mu}{d\lambda} + s\frac{d\theta}{d\lambda}. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

If $d\theta_1$ is the change in θ in passing from the order n to the order $n + 1$,

$$(n + 1)\lambda = (\mu - 1)t + s(\theta + d\theta_1),$$

and on subtracting the second last equation from this we obtain

$$\lambda = sd\theta_1.$$

If we substitute $sd\theta_1$ for $sd\theta$ in (3) and then equate $sd\theta_1$ to λ, (3) becomes

$$n = t\frac{d\mu}{d\lambda} + \frac{\lambda}{d\lambda}.$$

But n is approximately $(\mu - 1)t/\lambda$. Hence

$$\frac{(\mu - 1)t}{\lambda} = t\frac{d\mu}{d\lambda} + \frac{\lambda}{d\lambda},$$

or

$$d\lambda = \frac{1}{\frac{t}{\lambda}\left\{\frac{(\mu - 1)}{\lambda} - \frac{d\mu}{d\lambda}\right\}} = \frac{\lambda^2}{t\left\{\mu - 1 - \lambda\frac{d\mu}{d\lambda}\right\}}.$$

This expression gives the difference in wave-length for an angle equal to the angle between the images of successive orders. Its value can be calculated for all parts of the spectrum from the constants supplied by the maker, and the difference in wave-length corresponding to any other angle obtained by simple proportion.

Michelson's Interferometer. This instrument has already been described on p. 149, and it was applied by Michelson to the study of the structure of spectral lines before the invention of the echelon grating. Fig. 240 shows how it would be used for this purpose. A spectrum is formed of the source of light by an ordinary spectroscope. This spectrum falls upon a slit S which is placed so that only the line under investigation passes through it to the interferometer. The light from the slit is then rendered parallel by the lens L, and passes through the interferometer, and the bands are examined by the reading telescope T. The lens L would not be necessary if the source of light were a broad one like a sodium flame.

If the line under investigation is a doublet each component of the doublet produces its own system of fringes. If the mirror C is moved slowly out, the path difference of the two beams alters. For one position the systems superimpose and the fringes are very distinct; for another position the maxima of the one system fall on the minima of the other and the fringes cannot be seen. As C is moved out, the visibility of the fringes thus undergoes a periodic change, and from the travel of C the difference in wave-length of the two components can be calculated. If the line is a single one, there is no periodic change in the visibility of the bands. If the line contains more than two components there are more than two systems of bands, and the change in visibility is more complicated.

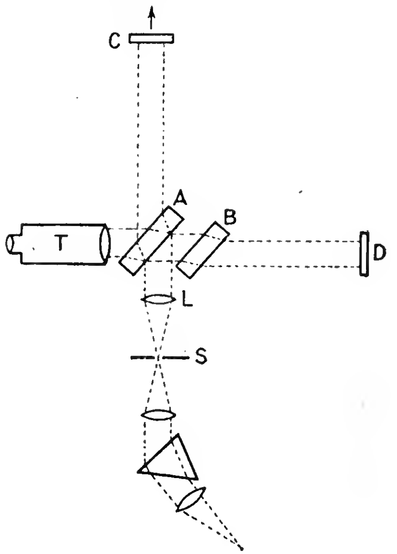


FIG. 240.

When used in this way the interferometer is like a diffraction grating with two rulings, one corresponding to each beam, and the resolving power is obtained by dividing the path difference by the wave-length.

Fabry and Perot's Interferometer. This apparatus consists essentially of an air space bounded by two parallel half-silvered glass surfaces. A half-silvered glass surface is one that lets through as much light as it reflects. Rays from a source P fall on this air space and suffer multiple reflection between the silvered surfaces. Consequently to an eye on the further side they

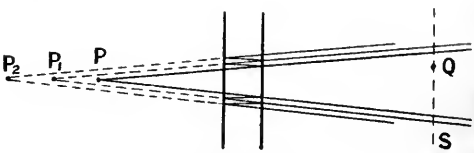


FIG. 241.

appear to come from a train of sources P, P_1, P_2, etc., the first of which is brightest with the others gradually decreasing in intensity. If we consider a point Q on a screen S near the axis of the figure, the length of the paths to Q of the beams coming from P, P_1, P_2, etc., increases in equal steps. The sources will thus reinforce if these steps contain an integral number of wave-lengths. As Q moves up the screen, the length of the step decreases and the different beams interfere, then when the step again becomes equal to an integral number of wave-lengths they reinforce one another again. Thus the screen is covered with circular interference fringes. As in this case there are a large number of interfering beams, the fringes are very sharp and the dark spaces between the fringes are much broader than the fringes themselves. Consequently when the spectral line used as source is not single but contains different components, and each of these components produces its own system of fringes, the different systems are clearly seen superimposed on one another, and from the distances between coincidences the ratio of the wave-lengths can be calculated.

Fig. 242 represents a Fabry and Perot interferometer; the two

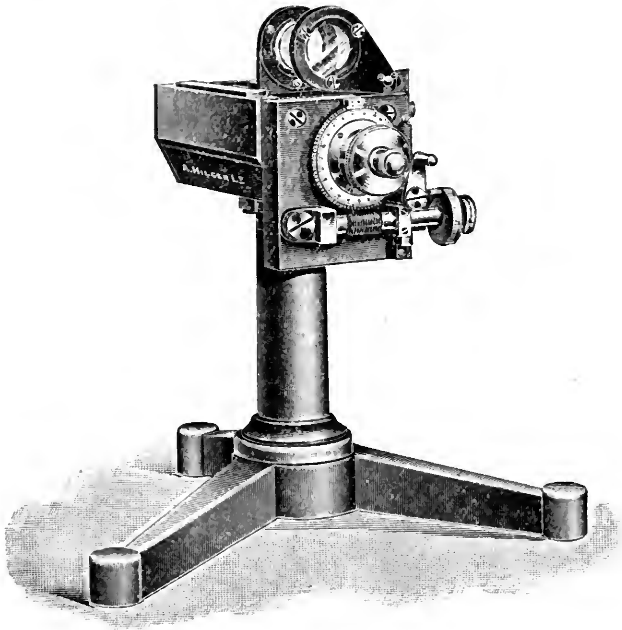


Fig. 242.

glass plates are clearly seen, also the arrangement for altering the distance between them. The latter can be increased up to 75 mm.

Lummer and Gehrcke's Interferometer. This consists of a parallel plate to which a small right-angled prism is cemented. The light

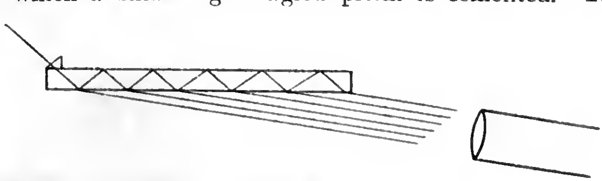


Fig. 243.

enters through this prism and is reflected back and forward inside the plate; at every reflection a portion of the beam is refracted out into the air. It is obvious that every two successive refracted beams have the same phase difference. The angle of incidence inside the glass is near the limiting angle of total reflection and consequently the intensity of the successive refracted beams diminishes slowly. Hence when they are received by a telescope they are in a condition to produce very sharp interference fringes.

EXAMPLES.

(1) Determine the value of the constant A in Balmer's formula from the experimental values of the wave-lengths for which $n = 4$, 6, and 10. Use your result to calculate all the terms of the series given on p. 272, and compare the calculated with the observed values.

(2) An observer and a source which emits monochromatic radiation are moving towards one another with velocities of v and u respectively. What effect has the relative motion on the wave-length as measured by the observer?

(3) Would the canal ray effect be visible in a spectroscope with a dispersing system of two large flint prisms?

(4) Prove that equations (2) represent motion in an ellipse.

(5) If an electron is vibrating under an attraction towards a point proportional to its distance from that point as represented by equations (1), what effect on the motion has the superposition of a constant electrostatic field?

(6) Look for the Swan spectrum in a bunsen burner and in an ordinary gas burner, and make a drawing of it.

CHAPTER XVII.

THE INFRA-RED AND X RAYS.

In 1800 Sir William Herschel moved a sensitive thermometer along a spectrum and found that the temperature reached a maximum at a point some distance beyond the red end. He thus proved the existence of heat rays, which did not excite the sensation of light. In 1830 Melloni made the first thermopile. If two wires of dissimilar metals, say copper and iron, are joined at both ends and a galvanometer is placed somewhere in the circuit, say in the middle of the copper wire, and if the one junction is kept at a constant temperature, for example, the temperature of melting ice, and the temperature of the other junction is greater than this, a current flows round the circuit. The direction of the current in the iron is from the hot to the cold junction, and the magnitude of the current is proportional to the difference of temperature between the two junctions, provided that the latter is not too large. If instead of one iron wire and one copper wire we take ten pieces of iron wire and ten pieces of copper wire, join them up alternately in a circuit and arrange them so that every second junction is in the melting ice and that the other junctions are raised to the former higher temperature, the electromotive force in the circuit is multiplied ten times. The different wires are, of course, insulated so that there is no short circuit anywhere and the current has to run round the whole circuit. Such an arrangement is called a thermopile. If every second junction is exposed to the heat rays and the other junctions sheltered from the heat rays and left to take up the temperature of the air of the room, the current in the circuit is a measure of the intensity of the heat rays. The magnitude of the electromotive force varies with the metals used for the wires ; Melloni employed antimony and bismuth, because they give a larger effect than copper and iron.

The thermopile is much more sensitive than the thermometer, and Melloni employed it in a number of investigations. He found what percentage of the incident dark heat rays was transmitted by various solids. Of all substances investigated rock salt was most transparent to the heat rays. The percentage transmitted by other substances varied with the temperature of the source, but in the case of rock salt it was always the same. Melloni came to the conclusion that the dark heat rays and the light rays were of exactly the same nature, that light rays were merely a kind of heat ray that had the property of exciting the retina. This view is of course the correct one, but it was some time before it was universally accepted. Many experimenters believed that the light rays were of quite another nature from the heat rays and existed in addition to them.

Heat rays, then, can be reflected, refracted and polarised in exactly the same way as light rays. Their reflection can be shown very easily to a large audience by means of concave metal mirrors. All metal mirrors reflect heat rays well even when they appear dull to the eye. If two concave spherical or preferably parabolic metal mirrors are placed some distance apart and facing one another and a copper sphere is heated in a bunsen flame to a dark heat and placed in the focus of the one mirror, there is a rapid rise of temperature at the focus of the other mirror. This can be measured with an ordinary thermometer. It will be noticed that if the thermometer is moved out of the focus there is no effect. The percentage of energy transmitted by a glass plate can be found by merely placing it between the two concave mirrors and noting the change in the reading; also the law of reflection can be verified by reflecting the rays at a metal mirror as shown by fig. 244.

Of course an ordinary thermometer would not make the effect visible to a large audience, and so in that case a thermopile and galvanometer or a differential air thermometer must be used. An air thermometer consists of two glass bulbs connected by a thin U tube, in which there is a coloured liquid. Alongside one vertical branch of the tube there is a

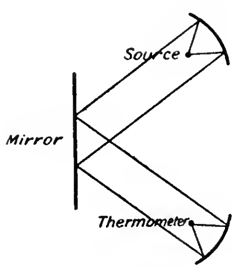


FIG. 244.

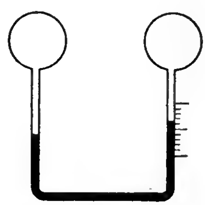


FIG. 245.

paper scale. If the glass bulbs were clear some of the heat rays would pass through; the bulbs are therefore blackened in order to absorb these rays and make all their heat go to raising the temperature of the air in the bulb. The instrument is placed with one bulb in the focus of the mirror and the other out of the way of the rays to the side. When the rays fall on the bulb, the air in it expands, the air in the other bulb contracts, and the end of the column of the coloured liquid moves up or down the scale.

In connection with experiments on heat rays perhaps it will not be superfluous to add a word of caution about error from extraneous heat sources. The observer must be careful of the effect of the heat of his own body, also there must be no glow lamps or rheostats near the thermopile.

The thermopile as a means of detecting heat rays was greatly improved by Rubens. In his instrument the wires are of iron and constantan and have diameters of from 0·1 mm. to 0·15 mm. Constantan is an alloy consisting of 60 per cent copper and 40 per cent nickel. There are 20 junctions for receiving the rays; these are soldered with beads of silver which are flattened into discs of about 1 mm. diameter so as to present a large surface to the rays. These 20 junctions are arranged along a straight line on a length of 2 cm., so that, when they are moved along a spectrum in a direction at right angles to this line, the radiation falling on them is approximately monochromatic. The chief improvement introduced by Rubens was in making the wires so fine; he thus diminished the heat capacity and enabled a steady deflection to be reached in a much shorter time.

In the visible spectrum the thermopile is not nearly so sensitive as the eye. In the case of the spectrum of a bunsen sodium flame, for example, the energy in the D lines is quite insufficient to produce a measurable deflection. Consequently the galvanometer used must be as sensitive as possible. For modern work two different types of galvanometer have been used with the thermopile, namely, the Du Bois-Rubens ironclad galvanometer and the Paschen galvanometer. In the former the suspension system consists of six very small magnets arranged parallel and close together with their poles pointing in the same direction. When the current flows through the coils these magnets are deflected. Both the coils and magnets are surrounded by two iron shells and an iron cylinder; the magnets are attached to a mirror by a light rod which passes through a hole in the shells. When the magnets turn, the mirror rotates and this rotation is read in the usual way by the movement of a spot of light on a scale. The purpose of the shells and cylinder is to screen off the earth's field and all disturbing magnetic fields and make the control field, i.e. the restoring couple, as small as possible. To prevent vibration of the laboratory from disturbing the galvanometer, it is usually suspended in a sling from a bracket on the wall. Such a galvanometer in conjunction with a good thermopile will in the most favourable circumstances measure an increase of temperature of 0·000001° C.

§ In addition to the thermopile there are three other instruments used for measuring the energy carried by a radiation. These are the bolometer, the radiomicrometer, and the radiometer. The bolometer depends on the principle, that if the temperature of a wire is raised, its resistance increases. The radiation is allowed to fall on a thin wire, the surface of which is blackened so as to make the reflection loss as small as possible. The wire forms one of the arms of a Wheatstone bridge, and the alteration of its resistance measures the intensity of the radiation. The bolometer was invented and its sensitiveness was brought to a very high pitch by Langley. It has been used more extensively than the thermopile, but it is not so easy to use as the latter.

The radiomicrometer, which was introduced by C. V. Boys, is a combination of galvanometer and thermopile. It uses the principle of the moving coil galvanometer. A little coil is suspended by a quartz fibre between the poles of a stationary magnet; below the coil and in the same circuit with it is a thermojunction, the elements of which are antimony and bismuth. When the rays fall on the junction a current flows round the coil and it turns in the magnetic field, the rotation being proportional to the intensity of the rays. A mirror is attached to the moving system and the magnitude of the rotation measured by the motion of a spot of light on a scale.

The radiometer as invented by Sir William Crookes consisted of mica vanes mounted on a central spindle after the manner of the blades of a paddle wheel. The arrangement was mounted in an exhausted tube, and when a beam of light fell on the vanes they rotated. This was due to the pressure of the gas left in the tube. The rays warmed the surface of the vane on which they fell, and the gas on that side had its temperature and consequently its pressure raised, while the temperature and pressure of the gas on the other side remained constant. Thus there was a resultant thrust on the vane in the direction of the rays, and as long as the rays fell on the vanes they kept spinning round. Instead of fixing the vanes to a spindle E. F. Nichols suspended two by a quartz fibre so that they were free to turn about a vertical axis, and allowed the beam of light to fall on one of them. The system consequently rotated until the action of the rays was balanced by the torsion in the fibre. A mirror was attached to the system, and the rotation was measured in the usual way by the excursion of a spot of light on a scale. The rotation was found to be proportional to the intensity of the rays. The instrument in this form is thus admirably adapted for the measurement of radiation, and it has been widely used for this purpose, especially in America.

§ The spectroscopes used for work in the infra-red have as a rule only a single prism, and instead of lenses they employ concave spherical silver mirrors. The material used for the prism is rock salt, sylvin, or fluorite; rock salt is cheaper than fluorite, but like sylvin it is hygroscopic, and if not looked after carefully requires frequent repolishing. Quartz is better than glass but not so good as the other materials. Glass transmits approximately to $2{\cdot}5\mu$, quartz to 4μ, while fluorite transmits to 11μ and rock salt and sylvin to 18μ.* The chief advantage of mirrors over lenses is that they do not require focussing. It is practically impossible to focus an infra-red spectrum; moving the thermopile through the spectrum tells us only about the position and intensity of a line but little about its sharpness; besides most of the spectra investigated are continuous ones. But if the mirrors are

* In work in the infra-red wave-lengths are always measured in μ. $1\mu = 10^{-4}$ cms.

focussed for the D lines, the spectrum will be in focus to the farthest infra-red. Another advantage of mirrors is their cheapness and the fact that they can be used into the farthest infra-red, where even fluorite and rock salt absorb.

A large number of the spectroscopes used in the infra-red are constant deviation ones. The reason for this is, that radiometers and radiomicrometers must remain stationary; they cannot be moved along a spectrum, the different radiations of the spectrum must on the contrary be allowed to fall on them in succession. Also the condition of constant deviation ensures at the same time the condition of minimum deviation and consequently maximum definition.

This will be made clearer by the description of a spectroscope which I have made and used in the infra-red, and which can be simply and inexpensively made by any amateur. This instrument uses a special case of the Wadsworth mirror-prism combination. The latter consists of a prism and mirror mounted together on a rotating table with the plane of the mirror and the plane that bisects the refracting angle of the prism intersecting in the axis of rotation of the table. It has the property that the rays which pass through the prism at minimum deviation and then fall on the mirror suffer a constant deviation.

The special case of the combination used is shown in fig. 246. CDB

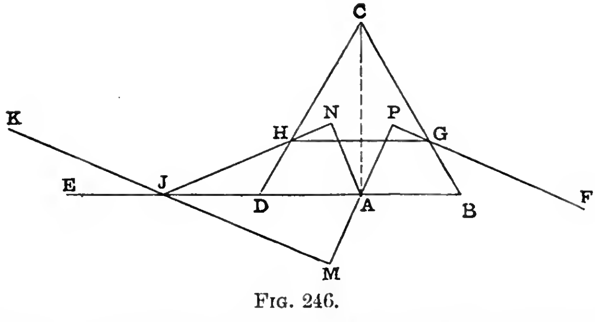


Fig. 246.

is the prism, ED the mirror, and A the point through which the axis of rotation passes. FGHJK is a ray which passes through the prism at minimum deviation and is reflected by the mirror at J. The path of the beam through the prism GH is consequently parallel to the base of the prism BD and KJ is parallel to FG. From A draw AP, AN, and AM respectively perpendicular to FG, JH, and KJ. Then AN = AM by equal triangles and AN = AP by symmetry; consequently AP = AM. If the ray FG is white light, the constituent colour that suffers minimum deviation emerges along JK after passing through the system. If the system is rotated through an angle about A and the ray FG remains fixed, AM remains fixed and the position of JK is unaltered. But it is now a different constituent colour that suffers minimum deviation and emerges along JK.

Fig. 247 shows how this property is taken advantage of in the construction of the spectroscope. S is the slit. The light diverges from S, falls on the concave mirror M, is then made parallel and passes through the mirror prism combination. After emerging it falls on the mirror M′ and is brought to a focus at T. If the thermopile is placed at T and the prism table rotated, the different colours in succession pass across T and each colour comes into minimum deviation

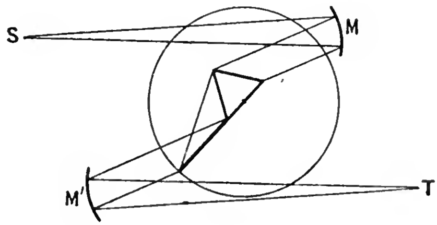


FIG. 247.

as it reaches T. This apparatus can, of course, be used also as a monochromatic illuminator in the visible spectrum. If a slit is placed at T and the prism system is rotated we have in succession monochromatic light of different wave-lengths emerging from this slit in a constant direction. Owing to the obliquity of the incidence the distances of the foci from their respective mirrors are given not by $\frac{1}{2}r$ but by $\frac{1}{2}r \cos \phi$, where ϕ is the angle which the principal ray of the beam makes with the normal to the mirror.

The mirrors and prism are best mounted in a box with openings at S and T. This keeps off stray light; also if a dehydrating agent is kept in it and the openings closed when not in use, it helps to preserve the prism if made of rock salt. Silvered glass surfaces are used for the mirrors. The outside surface of the silver is used. As it is somewhat difficult for an amateur to get a good surface on this side of the mirror, it is better to get the silvering done by an optical firm.

Diffraction gratings have not been used in the infra-red owing to the overlapping of their spectra, also owing to the fact that it is important for the spectra to be as intense as possible and their spectra are fainter than prism spectra.

Calibration of a Spectroscope in the Infra-Red. The indices of refraction of rock salt and fluorite are now known for the infra-red, and consequently the easiest method of obtaining the wave-length corresponding to a given position of the prism is simply by calculation. Or the scale may be calibrated by the use of known spectra. The absorption spectrum of water has been very thoroughly investigated by E. Aschkinass and is suitable for this purpose. It has well-marked absorption bands at 0·996, 1·500, 1·956, 3·02, 4·70, and 6·09μ. The band at 0·996μ shows up well when a layer of water 1 cm. thick is examined; the others require much thinner layers, $\frac{1}{20}$ mm. and less.

A straight Nernst filament is the best source to have as the background to this absorption spectrum. If a lens is used to project it on the slit, the image exactly fits the shape of the slit, whereas with the crater of an arc much light is wasted.

These methods were, of course, not available to the pioneer workers in the field. They used two methods. The first method was similar to Edser and Butler's method of calibrating the visible spectrum ; it consisted in producing interference bands across a continuous spectrum and mapping their positions in the visible spectrum with the eye, and in the infra-red with the thermopile or similar instrument. The visible spectrum was previously calibrated by other means, and consequently the wave-lengths of the positions of the interference bands in it were known. Thus the path difference and the order of the bands could be determined, and hence their wave-lengths calculated for the infra-red. To produce the bands H. Becquerel caused the light from the source before entering the slit to be reflected by an "air plate," i.e. a thin film of air bounded by two plates of transparent material. The interference took place between the rays which were reflected from the two faces of the air plate.

The second method was used and developed principally by Langley. It was first described in 1884. A spectrum of the sun was produced with a concave grating, and this spectrum was allowed to fall on the slit of the spectroscope to be calibrated. Suppose the D_2 line of the third order spectrum fell on the slit. Its wave-length is $\cdot 589\mu$; consequently superimposed on it was the wave-length $\cdot 883\mu$ of the second order spectrum and the wave-length $1\cdot 767\mu$ of the first order spectrum. The telescope which carried the bolometer was then moved round until the deviations of these lines were found. By taking different points in the grating spectrum and proceeding in this way the deviations were obtained for other known wave-lengths, and so a calibration curve could be constructed for the spectroscope.

§ Fig. 248 is an example of the best kind of results obtained in the infra-red.* It gives data obtained by W. W. Coblentz and W. C. Geer, and represents the emission spectrum of the mercury arc. The ordinates are radiometer deflections, the abscissæ are wave-lengths. The sharp maximum at $\cdot 54\mu$ is, of course, the bright green line. The maximum at $\cdot 58\mu$ is the two yellow lines, which were not resolved. In the region $0\cdot 9\mu$ to $2\cdot 1\mu$ three curves taken with different slit widths are given and each point on a curve is the mean of several readings. In the region beyond 4μ individual readings are given and two slit widths were worked with. The prism was of rock salt. It will be noticed that the deflections are very small ; also the curve gives little detail. With the same apparatus an acetylene flame gave a deflection of 50 cm. at the maximum part of its spectrum.

By bathing ordinary dry plates in solutions of certain dyes it is

* "Phys. Rev." **16**, 1903, p. 281.

possible to photograph the infra-red as far as 9000 or 10,000 A.U. By making a special silver bromide emulsion, which was blue by

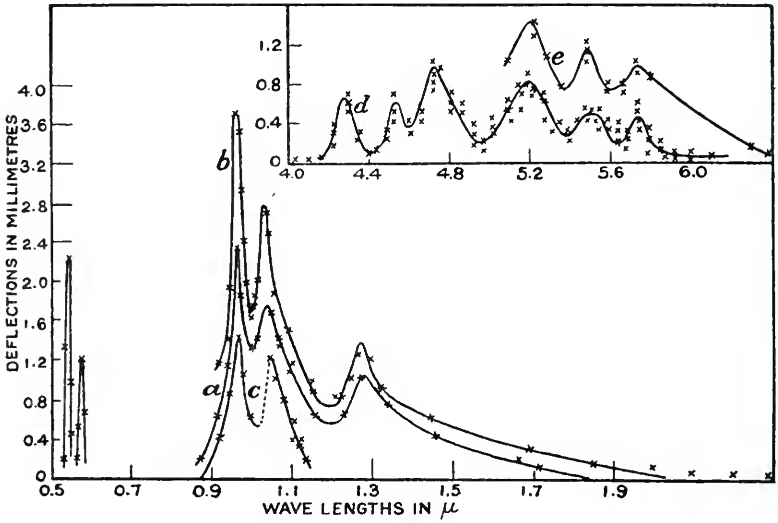


FIG. 248.

transmitted light instead of red as the ordinary emulsion is, Abney was able to prepare plates which were sensitive as far as 20,000 A.U. Such plates of course give far more detail than is obtainable with a thermopile, but as they are difficult to prepare and soon lose their sensitiveness they have not been used by succeeding workers.

§ The infra-red methods possess one advantage over visual observation and photography, namely, the deflection is proportional to the intensity of the spectrum. In the visible spectrum we can easily tell when one line is brighter than another, and we can also easily say which of two lines on a photographic negative is the stronger, but we cannot state definitely that the one line contains so many times more energy than the other. But if both lines fall in succession on a thermopile, and if the receiving surface of the latter is broader than they are, then the ratio of the deflections is proportional to the ratio of the energy in the two lines. The receiving surfaces of thermopiles, bolometers, etc., are always blackened so that the energy of all radiations is equally absorbed and changed into heat, no matter what their wave-lengths are. Thus the infra-red methods are specially adapted for measuring the variation with the wave-length of the intensity of a continuous spectrum, and much important work has been done in this direction.

Suppose, for example, that the continuous spectrum is received on a screen and we fix our attention upon that portion of it bounded by the wave-lengths λ and $\lambda + d\lambda$. The rays, the wave-lengths of which lie between λ and $\lambda + d\lambda$, bring every second a certain amount of energy to the screen; let this amount of energy be denoted by $E_\lambda d\lambda$. Then E_λ is a function of λ.

In mapping an energy curve, i.e. in determining the function E_λ, certain precautions and corrections must be attended to. First of all care must be taken that no stray or diffuse heat gets to the receiving surface. Figs. 249 and 250 illustrate what is meant. In fig. 249 the ther-

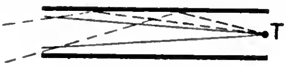


FIG. 249.

FIG. 250.

mopile T is mounted at the end of a tube. The full lines represent the pencil of rays falling on T, the intensity of which is to be measured. The dotted lines show another pencil of rays which have a different wave-length and leave the prism in another direction, but are reflected from the side of the tube and fall also on the thermopile. Now this second pencil may be much more intense than the first, and even a dead black surface reflects well at grazing incidence; consequently this may be a serious source of error. Fig. 250 shows how by mounting screens in front of the thermopile the error may be entirely eliminated.

The usual way of taking readings is to have a screen in front of the slit and set the spectroscope for the required radiation, then the observer watches the spot of light on the scale, and, when it is steady, removes the screen by pulling a cord or by some other arrangement. The spot of light at once makes a deflection on the scale and this deflection is noted. When the deflections have been taken for different points in the spectrum they may be plotted against the wave-lengths as abscissæ, but the curve obtained is not the energy curve. What is known as the "slit width correction" must first be made. Owing to the fact that the dispersion of the prism is not normal, the range of wave-lengths falling on the thermopile at different points in the spectrum, i.e. $d\lambda$, is not always the same. Since the deflections are proportional to $E_\lambda d\lambda$, in order to obtain the variation of E_λ we must allow for the variation of $d\lambda$. Let s denote the reading on the divided circle of the spectroscope, and plot s against λ, i.e. draw the calibration curve of the instrument. Determine $\frac{ds}{d\lambda}$ graphically from this curve. Then, if for any point on the spectrum the deflection is multiplied by the value of $\frac{ds}{d\lambda}$ for that point, the result is proportional to $E_\lambda ds$. Consequently, as ds is always the same, as the receiving surface has always the same width, it is this result that is taken as the ordinate of the energy curve. Of course in the region of the spectrum where the material of the prism absorbs, correction must be made for loss of energy from this cause.

The first energy curve studied was that of the sun, which Langley investigated several times with bolometers. His papers on the subject date from 1883 to 1900. He found that the solar spectrum could be followed to $\lambda = 18\mu$, although it became faint after 5μ. The solar spectrum is very much brighter than that of any terrestrial source, and so he was able to work with an extremely narrow receiving surface and obtained more detail than has ever been obtained by this method. In one of his arrangements the rock-salt prism was turned by clockwork and at the same time the deflections of the galvanometer were recorded by a moving light spot on photographic paper; the photographic paper was kept in motion by the same clockwork. This arrangement was sensitive enough to show the nickel line between the D lines in the solar spectrum. Langley was able to map the positions of 700 Fraunhofer lines in the infra-red.

Residual Rays. In 1896 E. F. Nichols working in Rubens' laboratory measured the reflecting power of quartz for various wave-lengths. The rays from a zircon burner were reflected from a polished quartz surface and then focussed by a rock-salt lens on the slit of the spectroscope. The latter had a fluorite prism. At 4μ, where it is transparent, the quartz was found to reflect 2 per cent of the incident light. At $8{\cdot}5\mu$ there was a maximum of reflection, 80 per cent of the incident light being reflected there.

This work was at once followed up by Rubens and Nichols. The rays from the heat source were made to suffer five reflections at polished quartz surfaces and then enter the slit of a grating spectroscope. The grating was made by winding wire round a frame. The spectrum was investigated with a radiometer. It was found that it consisted of a double maximum with its peaks at 8·5 and $9{\cdot}62\mu$ and another maximum at $20{\cdot}75\mu$. Quartz has metallic reflection at these three regions, or, in other words, it reflects these radiations as well as a polished metal surface reflects visible light. The source contained radiations of all possible wave-lengths, but the other wave-lengths were so much weakened at each successive reflection that they produced no deflection.

The monochromatic radiations produced in this way are called residual rays, and they have proved of great importance experimentally, because approximately monochromatic radiations of these wave-lengths cannot be produced by prisms owing to the material of the prism either absorbing them or not having a suitable dispersion. Fluorite has residual rays at $24{\cdot}0\mu$ and $31{\cdot}6\mu$, rock salt has residual rays at $52{\cdot}2\mu$ and sylvin at $61{\cdot}4\mu$.

Recently a new and interesting method has been used for the isolation of long heat waves. Rubens had found that quartz was transparent for the longest residual rays previously known and had an index of refraction of about 2·2 for these rays. As source of light a Welsbach incandescent mantle without a glass funnel was taken, the focal length of a quartz lens was calculated for this index of refraction, two screens, B and E, with holes in them were placed four times this distance apart, and the lens placed midway between them. Consequently the very longest heat rays from the source after diverging from the hole in B were brought to a focus by the lens in the hole in E. The shorter waves did not converge to the same extent, or in the case of the visible and ultra-violet rays even diverged slightly. Hence they fell on a wide area of the screen E, and in ordinary circumstances a small

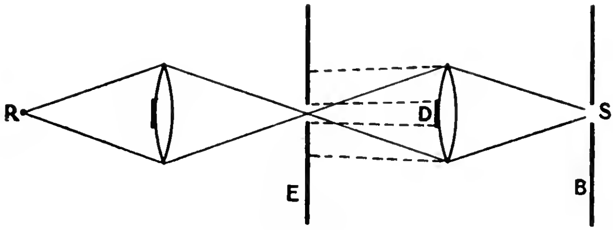


FIG. 251.

proportion of them would get through the hole. To prevent this a disc D was fastened to the centre of the lens with wax to block out these central rays. A second lens was used to focus the rays on the element of a radiomicrometer. It purified the rays further.

The wave-length of the rays isolated in this way has been found to be 107μ or more than one-tenth of a millimetre.

Phosphorescence. Certain substances, after being exposed to light, emit light for some time afterwards when placed in a dark room. This phenomenon is known as phosphorescence. It is the violet and ultra-violet light that are most active in producing it. The duration of the emission after exposure to light varies very widely. Balmain's luminous paint, which is a sulphide of calcium, will shine for hours in the dark after exposure to bright sunshine, while other substances cease to emit in a few seconds. To detect the phosphorescence in these cases Becquerel has invented an instrument called the phosphoroscope, which consists essentially of two discs on the same axle. These discs are each pierced with the same number of circular holes arranged on a circle at equal distances the whole way round, but they are mounted out of step, i.e. the holes on the one wheel are half-way between the holes on the other. The substance to be investigated is placed between the two discs, the exciting light comes from the one side through the holes in the one disc, while the observer looks through the holes in the other disc. The observer does not see the substance when the exciting light is falling on it but a short interval of time after, when it is cut off. How short this interval of time is depends on the speed at which the discs are rotating. When examined in this way all solid fluorescent substances are found to be phosphorescent.

The phenomenon is quite distinct from the faint greenish-white luminosity which phosphorus shows when exposed to moist air in the dark. The latter is due to the slow combustion of the phosphorus. It is a chemical change which takes place only in the one direction and is not reversed by exposure to light. In true phosphorescence we have a reversible change; the exciting light throws into the substance the energy which is radiated away in the dark, and by exposing the substance to the exciting light it can be made to phosphoresce again and again.

Effect of Infra-red Rays on Phosphorescence. If a card, which has been coated with Balmain's luminous paint, is made luminous by a short exposure to sunlight and then has an infra-red spectrum focussed on it, the parts on which the infra-red rays fall phosphoresce more brightly than the rest and then become exhausted, so that if the card is afterwards removed to a dark room they appear dark while the rest of the paint is still phosphorescing. Heat has exactly the same effect on the paint. This property of the rays of affecting phosphorescence has been used for mapping infra-red spectra.

Ångström's Pyrheliometer. The various instruments described on pp. 297-8, the thermopile, bolometer, radiomicrometer, and radiometer measure the relative intensities of two radiations but do not give the intensity of any one radiation in absolute measure, i.e. they do not give the strength in ergs/sq. cm. sec. Ångström has found that the Hefner lamp radiates $2{\cdot}15\ 10^{-5}$ calories per square centimetre per second at a distance of 1 metre in a horizontal direction. Hence by exposing any of the above instruments to the radiation from a Hefner lamp their readings can be standardised and converted to absolute measure.

Instruments called pyrheliometers have been invented for giving the intensity of any radiation, especially that of the sun, directly in absolute measure. The most celebrated of these, Ångström's pyrheliometer, consists essentially of two metal strips blackened on one side and in every way similar. One of these strips is exposed to the radiation to be measured, while the other, which is screened by a double wall from this radiation, is heated by an electric current. The strength of the current is regulated so that the temperature of the two strips is the same, as read by a thermocouple and galvanometer. Then the energy expended by the current in the one strip is equal to the energy radiated into the other. Let q be the intensity of the radiation in cals./sq. cm. sec., l, b the length and breadth of the strip, r its resistance per unit length, a the fraction of the incident radiation it absorbs, and i the strength of the compensation current. Then

$$qalb = \frac{lri^2}{4{\cdot}2},$$

which gives

$$q = \frac{ri^2}{4{\cdot}2\ ab}.$$

The values of the constants a, b, and r are determined once for all. The convection and radiation losses are the same for both strips and consequently do not require to be corrected for. The rôle of the strips is interchangeable. The strips are of platinum foil about 0·001 to 0·002 mm. thick; they are about 18 mm. long and 2 mm. wide.

Kathode Rays. If the electrodes of a vacuum tube are connected to the secondary of an induction coil and the pressure in the tube is gradually reduced, the appearance of the discharge changes in a marked manner. At atmospheric pressure the tube either does not conduct or, if it does, the spark passes as a thin bright line between the electrodes. As the pressure is decreased, this line appears, if it is not already there, then widens out and fills the whole tube with a diffuse glow known as the positive column. The tube at the same time conducts much better and the discharge becomes smoother. When the pressure reaches 8 or 10 mm. of mercury the kathode is covered with a soft glow, the negative glow, and between the negative glow and the positive column a dark space, the Faraday dark space, appears. As the pressure is reduced to $\frac{1}{2}$ mm. of mercury, the conductivity of the tube begins to diminish, the anode becomes surrounded with a glow, and the positive column breaks up into a number of bright and dark spaces, striæ as they are called, as shown at E (fig. 252). At the same time the Faraday dark space D widens, the negative glow C detaches itself from the kathode, and a new luminosity forms on the surface of the kathode. Thus the negative glow is divided into two parts which are separated by a sharply defined dark space B, called the Crookes' dark space or the kathode dark space.

FIG. 252 (from Watson's "Physics").

At higher rarefactions both the positive and negative glows become less definite and luminous, the resistance of the tube increases, and the kathode dark space grows in size until finally at a pressure of $\frac{1}{1000}$ mm. of mercury it fills the tube. The sides of the tube shine then with a brilliant fluorescence, which is emerald green in the case of soda glass and blue in the case of lead glass. This fluorescence is produced by rays emitted from the kathode impinging on the glass. The rays proceed in straight lines from the kathode, for if a screen is placed between the kathode and the walls of the tube, a sharp shadow of the screen is produced on the walls. When the tube is placed in a magnetic field, the rays are deflected in the same way as a flexible wire carrying a current from the anode to the kathode would be deflected. They are also deflected by an electric field in the same way as a negative charge would be deflected. By a measurement of the magnetic and electric deflections it has been shown that the rays consist of electrons and that their velocities range from 10^9 to 10^{10} cms./sec., i.e. from one-thirtieth to one-third of the velocity of light. When the rays strike an obstacle most of their kinetic energy is dissipated as heat.

X-Rays. It was discovered by Prof. W. K. Röntgen at Würzburg in 1895 that when kathode rays encounter matter, it not only fluoresces and rises in temperature but also emits an entirely new kind of radiation, Röntgen radiation or X-rays. These rays make many substances fluoresce, they travel in straight lines, and they have also the striking property of penetrating media that are opaque to ordinary light. They pass through paper, cloth, leather, and aluminium, but are stopped by the more dense metals, especially lead. Flesh is transparent to them but the bones are opaque, and by their aid we can see the bones through the flesh. They thus have an important application in medicine.

Fig. 253 shows diagrammatically a tube used for producing X-rays.

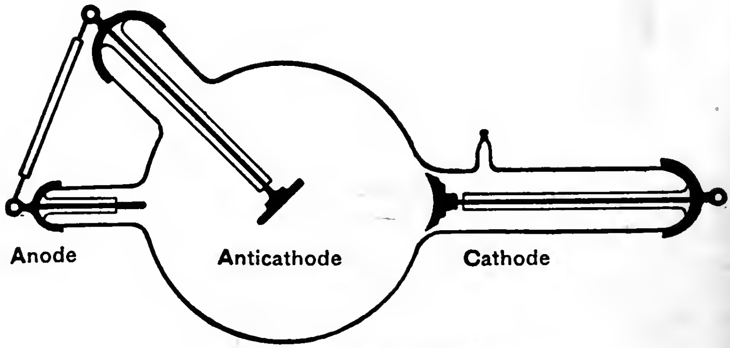


Fig. 253 (from Kaye's "X-Rays").

The kathode is concave and concentrates the kathode rays on the anti-kathode. The X-rays are radiated out from the anti-kathode and pass through the glass wall of the bulb undeflected. The fluorescence of the latter is caused by reflected kathode rays, not by X-rays passing through it. The anode is not essential to the working of the bulb but is supposed to steady the discharge. It is in conducting communication with the anti-kathode. The discharge is also made steadier by having the leading-in wires for the electrodes sheathed in glass tubes as shown in the figure. The kathode and anode are usually made of aluminium. The anti-kathode is made of a metal such as platinum with a high atomic weight and high melting-point. In many bulbs there is an arrangement for preventing the anti-kathode from becoming too hot.

X-rays produce fluorescence when they fall upon a screen of barium platino-cyanide. They also act on a photographic plate, and on these two properties their practical application depends. If the object is to be examined visually it is placed between the bulb and a fluorescent screen, and the radiograph or shadow picture produced is examined from the other side. Simple objects used for demonstration purposes are a purse with coins in it, a wooden box containing a set of brass weights for a balance, a resistance box, etc. The purse comes out transparent but the coins in it appear opaque. The sleeve of the arm holding the purse is transparent but the sleeve-links or buttons are opaque. The brass weights in the box and the resistance coils are both visible through their wooden covers. If a permanent record of the radiograph is desired, a photographic plate is put in the place of the fluorescent screen. Fig. 254 gives a radiograph of a hand showing rings and bracelet.

Further Properties of X-Rays. When an X-ray bulb has been run some time, it is said to become "hard," that is, its resistance increases and a higher voltage is required to run the bulb. If E is the potential difference to which the kathode ray owes its velocity v, then

$$\tfrac{1}{2}mv^2 = \mathrm{E}e,$$

where e and m are the charge and mass of an electron. If e/m is put $= 1{\cdot}77 \times 10^7$ and E is taken in volts and v in cms./sec., then

$$\mathrm{E} = 2{\cdot}82\, v^2\, 10^{-16},$$

that is,

$$v = 5{\cdot}95 \sqrt{\mathrm{E}}\, 10^7.$$

Thus as the bulb becomes harder, the velocity of the kathode rays increases.

At the same time the X-rays become more penetrating. It is customary to specify the penetrating power of rays by their absorption in a sheet of aluminium of definite thickness. If I_0 is their intensity before entering and I their intensity after passing through a plate of thickness d cms., it is found experimentally that for homogeneous X-rays

$$\mathrm{I} = \mathrm{I}_0 e^{-\lambda d},$$

where λ is a constant for the medium. This is the same law as holds for the absorption of light (cf. p. 328). λ is termed the linear absorption coefficient of the rays. The smaller λ is, the more penetrating the ray.

According to Eve and Day the value of λ for "soft" rays in air, i.e. rays emitted from a soft bulb, is from 0·0010 to 0·0018, and for "hard" rays is about 0·00029.

It is found that λ is not constant for the rays given out by an ordinary bulb, being greater for thin screens than for thick ones. This is due to the fact that such rays are not homogeneous but contain radiations with different values of λ, and the radiations with the greater values are naturally absorbed first.

Only a very small proportion of the energy of the kathode rays reappears as X-rays, perhaps about $\frac{1}{320}$ in the case of a platinum antikathode in a bulb of medium hardness.

When X-rays enter a gas they ionise it, that is, negatively and positively electrified particles are formed, and the gas becomes able to conduct electricity. The view generally accepted of the formation of these ions is that an electron is removed from the molecule, leaving the latter with a positive charge. Both the electron and charged molecule then become nuclei of clusters of gas molecules, though at low pressures the electron is able to exist alone.

The intensity of the X-rays can be measured by the number of ions they produce. One way of doing this is by means of an electroscope; the rays are sent directly into the electroscope, the gold leaf is charged to a high potential and the degree of ionisation measured by the rate of leak to the outer case.

Characteristic X-Rays. When X-rays strike a substance, it gives off in general scattered X-rays, characteristic X-rays, and electrons. The scattered X-rays are identical with the exciting rays in quality and may be compared with the ultra-violet light scattered by a grain of fluorescein. The characteristic rays, on the other hand, are homogeneous in quality and may be compared to the fluorescent green light sent out by the same grain. In order to produce the characteristic rays, the exciting rays must be harder than they are themselves. Many elements give out at least two characteristic radiations, and Barkla has called these the K and L radiations. For each metal the K radiation is about 300 times more penetrating than the L radiation.

The characteristic rays are also emitted by the antikathode in an X-ray bulb, especially if the tube is a soft one.

The Diffraction of X-Rays. For a long time there was considerable doubt as to the nature of X-rays. They were neither refracted nor reflected; also experiments with a V-shaped slit did not give definite proof of their being diffracted. We know now that they are extremely short light waves, and that the early experiments failed to give a positive result on account of this shortness.

A surface reflects and refracts light regularly only if the inequalities in it are small in comparison with the wave-length of the light, and a diffraction grating forms spectra only if the irregularities in the rulings are small in comparison with the wave-length of the light. Now if

FIG. 254 (from Watson's "Physics").

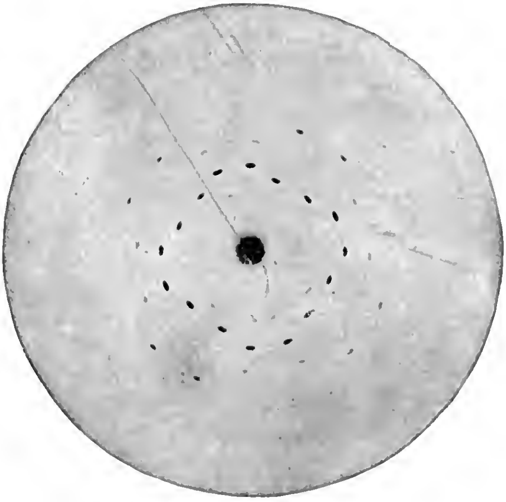


FIG. 255 (from Kaye's "X-Rays").

[*To face page* 310.

the wave-length of the light in question is $\frac{1}{1000}$ of the wave-length of Na light, it will be of the same order of magnitude as the diameter of the molecules, consequently it will be impossible for us by any means whatever to prepare an artificial surface smooth enough to reflect and refract waves of this length, or to construct a grating with the rulings close enough together to form diffraction images with light of this wave-length. But according to modern crystallography atoms in a crystal are regarded as forming a perfectly regular system of points in space. The systems formed by the different atoms interpenetrate, the result being to form what is called a " space-lattice ". It occurred to M. Laue, that if X-rays were of extremely short wave-length the regular structure of the crystal itself might act as a diffraction grating; thus nature might supply us with a grating with a structure finer than can be made by artificial means. On Laue's suggestion Friedrich and Knipping tested this point experimentally. The X-rays from a bulb were cut down by lead stops so as to form a thin parallel pencil. This pencil then fell on the crystal, and a photographic plate was placed behind the latter at a distance of a few centimetres from it.

The result of the experiment was a brilliant verification of Laue's idea. The photographic plate showed an intense spot formed by the undeflected rays surrounded by a number of diffraction images as shown in fig. 255, which is taken from a joint paper by Friedrich and Knipping and Laue, and shows the system of spots formed by a zinc-blende crystal. The pattern on the plate was altered by rotating the crystal. Exposures of some hours were necessary to bring out all the spots, since much the greater proportion of the rays was undeviated. The calculation of the positions of the spots is not simple; unlike an ordinary diffraction grating the crystal does not give a continuous spectrum, even though the rays contain a continuous range of wave-lengths.

§ Following on Laue's work W. L. Bragg found that X-rays are regularly reflected by cleavage planes in crystals; such planes are rich in atoms and, of course, the atoms in them are arranged regularly. Also since the atoms are arranged in a series of planes parallel to the cleavage plane and equidistant from one another, when the rays fall on these planes, the crystal acts like a system of equidistant semi-transparent mirrors. If the phase difference of the beams reflected in any direction from the successive planes is equal to an even number of wave-lengths, there is a maximum of reflection in that direction.

An instrument utilising this property and known as the X-ray spectrometer has been used recently (1913) by both W. H. Bragg and W. L. Bragg and by Moseley and Darwin. In it the collimator is replaced by a tube with lead stops. A crystal is mounted on the spectrometer table to reflect the rays, and the telescope is replaced by an ionisation chamber or photographic plate. The angle of reflection of the rays is always equal to the angle of incidence, but if the waves reflected by the successive planes reinforce in any direction, there is a maximum of reflection in that direction. If the strength of the reflected beam is plotted as a function of the angle of incidence, the curve has sharp maxima. These correspond obviously to homogeneous radiations contained in the incident beam. Just as in the case of the diffraction grating, there are spectra of different order according as the path difference between the different waves is one or more wave-lengths. By making assumptions as to the structure of the crystals employed it has been found possible to determine with certainty the wave-lengths of the different homogeneous radiations. The greater part of the radiation from an ordinary X-ray bulb is analogous to white light and gives a continuous spectrum. Superimposed on this are the characteristic radiations, which are analogous to spectral lines. The K and L radiations are now known to be bands in the spectrum owing to the constant values obtained for their absorption coefficients, but with the X-ray spectrometer spectra can be mapped in much greater detail than by absorption methods. The wave-lengths of the homogeneous X-rays already determined vary from more than $3\ 10^{-8}$ to less than $1\ 10^{-8}$ cms. The hardness or penetrating power of an X-ray depends on its wave-length, the ray being harder the shorter the wave-length.

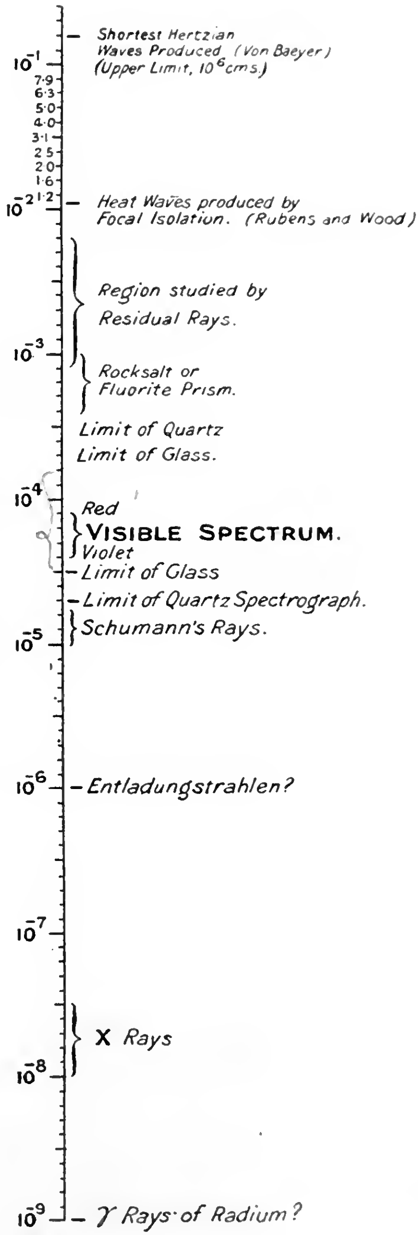
The γ rays of radium are of the same nature as X-rays. Their wave-length is possibly about 10^{-9} cms.

The discovery of the nature of X-rays has thus carried our knowledge of spectra an enormous distance into the ultra-violet. To show what a range of radiations we now have power over, all the different kinds with the exception of the longer waves used in wireless telegraphy have been plotted logarithmically against their wave-lengths on the opposite page.

EXAMPLES.

(1) Make an experimental determination of the fraction of the incident heat radiation transmitted by glass plates, mica plates, glass cells filled with water, etc., by letting the radiation fall upon a thermopile and then placing the substance to be investigated in its path. (The experimental details will depend on the resources of the laboratory, but it should be noted that a Nernst lamp with a single filament is much the best source for work of this kind.)

(2) A thermopile is placed in front of a Nernst filament and a wooden screen with a black surface placed between. When the screen is removed the galvanometer is deflected. The deflection is due not to the radiation from the filament but to the difference of the radiations from the filament and from the screen. How great is the error involved in ignoring the radiation from the screen? Assume that the thermopile is in an enclosure with black walls which are at a temperature of 15° C., and that the radiation enters through an aperture which subtends a solid angle of $\frac{1}{4}$ at the thermopile, that the filament subtends a solid angle of $\frac{1}{300,000}$ at the thermopile, that the screen



In the scale on this page the different kinds of rays known to us are arranged logarithmically against their wave-lengths. It will be seen that the range of the latter has grown enormously and now extends from 10^6 cms. to 10^{-9} cms.

Entladungstrahlen are produced when an electric spark passes in air but not much is known about them, the region between Schumann's Rays and X-Rays being practically unexplored.

The waves used in wireless telegraphy are, of course, long Hertzian waves.

is at 15° C., that behind the filament there is a black wall at 15° C., and that the radiation emitted from 1 sq. mm. of filament is 2500 times the radiation emitted from 1 sq. mm. of black surface at 15° C. (The solid angle subtended by a surface at a point is numerically equal to the area intercepted on a sphere of unit radius with the point as centre by the infinite number of lines which can be drawn from the point to the edge of the surface.)

CHAPTER XVIII.

PHOTOMETRY AND SPECTROPHOTOMETRY.

A POINT source of light radiates equally in all directions. Consequently if a sphere is drawn with the source as centre, the quantity of energy received per unit area is the same all over the surface of this sphere. The quantity of energy radiated through the surface of the sphere is always the same no matter what the area of the surface is. The quantity received per unit area must therefore vary inversely as the area of the surface, $4\pi r^2$, or, in other words, the illumination of a surface varies inversely as the square of its distance from the source. This relation, which is known as the inverse square law, enables us to compare the intensities of different sources, and thus forms the basis of photometry.

We shall now proceed to describe some of the principal photometers.

The Rumford or Shadow Photometer. This consists of a rod of the shape of a lead pencil, which is mounted vertically before a screen. Fig. 256 represents a plan of the arrangement; R is the rod, AC the screen, and L_1 and L_2 the two sources the intensities of which are to be compared. BC is the shadow of the rod formed by L_1 and AB the

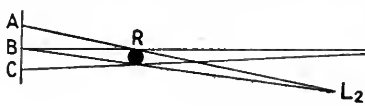


FIG. 256.

shadow of the rod formed by L_2 on the screen. The distances of L_1 and L_2 from the screen are adjusted so as to make these shadows equally dark, and at the same time R is placed so that the shadows just touch at B. If they overlap, or if there is a bright space between them, the comparison of their intensities cannot be made so accurately.

All parts of the screen outside the regions AB and BC are illuminated by both sources. The region AB being in the shadow of L_2 is illuminated solely by L_1, and the region BC being in the shadow of L_1 is illuminated solely by L_2. Consequently if the shadows are equally intense, the illuminations produced by the two sources at B are equally great, and the intensities of the two sources are simply as the squares of their distances from B.

The screen should have a rough unglazed surface so that nothing of the nature of regular reflection takes place; it may be made of white drawing paper or tissue paper or ground glass. In the former case it is observed from the front, i.e. from the side on which the light sources are situated. In the latter case it is observed from the back.

One great advantage of the shadow photometer is that the accuracy of its results is not affected much by the presence of another source of light in the room. For this third source produces a shadow of its own on the screen, which can easily be recognised, and it is only necessary to take care that the edge of this third shadow does not fall on B.

The Bunsen or Grease Spot Photometer. This consists of a piece of white unglazed paper in the middle of which a small portion has been rendered translucent by a drop of grease. It is placed between the two sources and in the same straight line with them. It possesses the advantage of simplicity, but it is difficult to make a grease spot with a very sharp boundary, and, of course, it is impossible to detect slight inequalities in the illumination when the areas are not sharply divided. As a substitute for the grease spot a disc of tissue paper may be used. The latter is made by placing two sheets of paper on the top of one another and with a sharp knife cutting a star-shaped opening in both of them at the same time. The tissue paper is placed between them, and they are then stuck together with the stars accurately superimposed.

The theory of the grease spot photometer is usually given as follows: When unit quantity of light falls on the grease spot, let the fraction b be diffusely reflected and let the fraction $1 - b$ be transmitted; when unit quantity of light falls on the white paper, let the fraction a be diffusely reflected and let the fraction $1 - a$ be transmitted. Let us suppose now that the distances of the sources have been adjusted so that they produce equal illuminations at the photometer. Then, if we look at the grease spot, the amount of light coming through it is proportional to $1 - b$, the amount of light reflected by it is proportional to b, and the total amount of light received from it proportional to $1 - b + b = 1$. Similarly the amount of light reflected by the white paper will be proportional to $1 - a + a = 1$. Consequently the grease spot will appear the same brightness as the surrounding paper. Thus to use the photometer we should adjust the distances until the grease spot disappears; then the intensities of the sources should be as the squares of their distances from the photometer.

In practice, however, it is found that the grease spot seldom disappears for the one position, when viewed in succession from both sides of the screen. This is partly due to the fact that the grease spot and white paper do not usually absorb the same fraction of the incident light, and absorption is entirely neglected in the preceding theory, but it is caused principally by the light reflected by the white paper being scattered more than the light transmitted by the grease spot. The relative intensity of the two surfaces depends thus on the angle at which they are viewed. The more obliquely they are viewed, the darker appears the grease spot.

Thus in general in working with the grease spot we set it so as to obtain the same relative intensity of the two areas when viewed from both sides, or in other words, we adjust, not for disappearance, but for the same contrast effect. For this purpose it is an advantage to be able to see both sides at once, not one after the other, and so two mirrors AM and AN are often used as shown in fig. 257; G is the grease spot and the dotted line gives the direction of the two sources to be compared. The eye E observes from the side, and the distances are adjusted until the images of the grease spot seen in the two mirrors show the same contrast against their respective backgrounds.

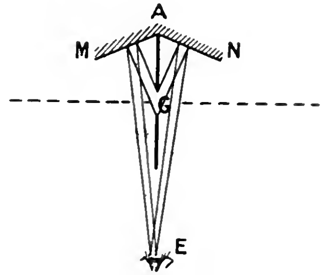


FIG. 257.

Stray light seriously affects the accuracy of the grease spot photometer, making the source on the side from which it falls appear too intense. This error can be eliminated by interchanging the positions of the two sources, the correction being analogous to the correction for unequal length of the arms of a balance by double-weighing.

The Wedge Photometer. The use of this photometer, which is sometimes called the Ritchie wedge, is shown in the adjoining fig. 258. S_1 and S_2 are the sources. The wedge is arranged so that its faces make equal angles with the line joining S_1S_2. The faces are rough and reflect the incident light diffusely. The observer looks at the edge of the wedge from the side at E, and adjusts the distances of the sources so that the two faces appear equally bright. The intensities of the sources are then as the squares of their distances from the edge.

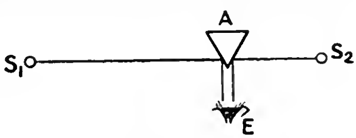


FIG. 258.

A scraped surface of plaster of Paris forms the best diffusing surface, but instead of this a wooden wedge may be used and a sharply folded piece of unglazed paper placed over it and fastened at the back A with a drawing pin. The sharper the fold, the more accurately can the setting be made. I have also found it an advantage to have a sighting arrangement attached to the wedge so as to make sure that the edge is viewed at right angles to the axis of the bench.

Joly's Diffusion Photometer. This consists of a rectangular block of paraffin wax cut into two equal blocks, which are then placed together again with a sheet of tinfoil between. The light from the sources enters into the blocks and is scattered inside them. Consequently they appear luminous when viewed by an eye situated on the same side as E. The positions of the sources are adjusted until both

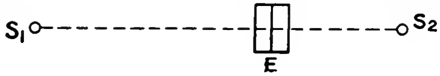


FIG. 259.

halves of the block appear equally luminous. Then their intensities are as the squares of their distances from the photometer.

The Lummer-Brodhun Photometer. This instrument employs the same principle as the grease spot photometer but in a much more refined form. Fig. 260 is a plan of the arrangement. A is a diffusing screen which is placed between the two sources, the direction of the latter being indicated by the dotted line. B and C are mirrors. P and Q are totally reflecting prisms, and E is an eyepiece which focusses on their hypotenuse surfaces. The hypotenuse surface of Q is plane and that of P is curved. Consequently they are in contact only over a small area. This area is in focus in the centre of the field of the eyepiece, and is illuminated by light diffused from the left-hand side of A and then reflected by B; the rest of the hypotenuse surface of Q is illuminated by light which is totally reflected on that surface and which comes from the right-hand side of A and is reflected at C. The area in contact is thus analogous to the grease spot, but unlike the latter it transmits all the light that falls on it and reflects none, while the surrounding part of the hypotenuse surface is analogous to the white paper surrounding the grease spot, but unlike the latter it reflects perfectly and does not transmit or absorb at all. Since a very thin air layer is sufficient to cause total reflection the area in contact is very sharply defined. As the two sources are moved along the dotted line to and from A, the relative intensity of the central spot and surrounding area alters, and, when the illumination of both sides of A is the same, the central spot disappears. The intensities of the two sources are then in the ratio of the squares of their distances from A.

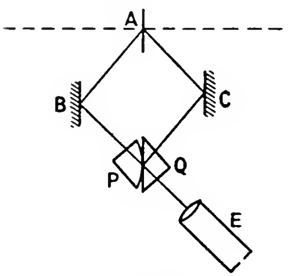


FIG. 260.

This photometer although usually accredited to Lummer and Brodhun was nevertheless invented very much earlier by Swan. But as there was no need for accurate photometry in his time, it was forgotten, and not brought into use until invented independently by Lummer and Brodhun.

In addition to the disappearance photometer Lummer and Brodhun have also invented a contrast photometer which is the most sensitive of all photometers. Fig. 261 shows the appearance in the field. The latter consists of two parts, each with a darker patch in its centre, and the distances of the two sources are adjusted until the two halves A and B have the same intensity and at the same time the patches *a* and *b* stand out in equal contrast to their backgrounds. Fig. 262 shows how

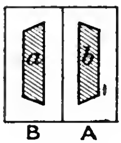


FIG. 261.

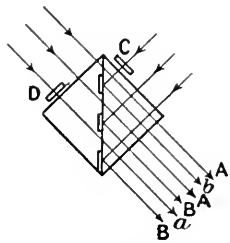


FIG. 262.

this result is attained. Instead of making contact at one spot the prisms make contact in the manner shown. The half A of the field and the patch *a* are illuminated by light from the one source, only the light from the patch *a* goes through a glass plate D before reaching the prisms. This has the effect of diminishing its intensity by about 8 per cent. The half B of the field and the patch *b* are illuminated by light from the other source, the light from the patch *b* being similarly weakened by passing through a glass plate before reaching the prisms. Thus, when balance is attained, the dividing line between A and B disappears and at the same time the two patches appear 8 per cent darker than their surroundings. The intensities of the sources are as the ratio of the squares of the distances from the diffusion screen just as in the case of the other Lummer-Brodhun photometer.

Photometry. Any of the photometers described in the preceding sections may be placed together with the light sources on the top of a table and the distances between them simply measured with metre sticks. But the results obtained in this way would not be accurate. The photometer and light sources are usually mounted on pieces which slide along a bar about 3 metres long. This bar carries a scale on which the distances between the pieces can be read. The bar and the sliding pieces should be blackened to prevent their reflecting light to the photometer. The ceiling of the rooms and the walls, particularly the parts behind the light sources, should also be blackened. If this is not possible, the error due to stray light can be eliminated by mounting diaphragms with dead-black surfaces on the photometer bar in such positions that they transmit all the direct rays from the source to the photometer but stop light reflected from the walls of the room.

The inverse square law holds rigorously only for point sources, hence care must be taken that the distance between source and photometer is always large in comparison with the linear dimensions of the source.

Generally speaking, in photometry an accuracy of 1 per cent in the final result is very good, but under the most favourable circumstances $\frac{1}{10}$ per cent can be attained. But in fitting up a photometer and taking readings for the first time the accuracy attained is always very much less than this. Some of the readings diverge widely from the mean, and I have found that the temptation to students to throw these away and use only the "good readings" is a very strong one. This should never be done. The "bad" readings keep the average straight. If a setting is made carelessly, the numbers should not be read, but once they are read and noted down on paper they should contribute their share to the final result, unless, of course, there is something obviously wrong about them.

A photometer bench can be used to measure the fraction of light transmitted by a glass plate. In order to do this two lamps are balanced against one another. Let the distance from the photometer of one of these lamps be d_1. Place the glass plate in the path of the rays from this lamp; owing to the loss due to absorption in the plate and reflection at its surfaces the lamp must be moved nearer the photometer in order to restore balance. Let d_2 be its distance from the latter when balance is restored. Then the fraction of the incident light transmitted by the plate is obviously $(d_2/d_1)^2$.

The Standard Candle. This is the historical unit of intensity of a light source. It has always been customary to express the intensity of a light source in candle-power. The British standard candle is a spermaceti candle, $\frac{7}{8}$ inch in diameter, weighing six to the pound, and burning at the rate of 120 grains per hour, but the brightness of a candle flame varies with the length and shape of the wick. It is thus very unsatisfactory as a standard, and consequently in modern British practice the unit of intensity is taken as one-tenth part of the intensity of the Vernon Harcourt 10 candle-power pentane lamp, burning at a pressure of 760 mms. mercury in an atmosphere containing 8 parts in 1000 by volume of water vapour. This lamp burns an inflammable gas formed of the mixture of pentane vapour and air. Pentane is a volatile hydrocarbon (C_5H_{12}). The mixture is formed in a carburetter and is supplied to an Argand burner. The flame is fed with preheated air.

Violle suggested as a standard the light emitted normally by one square centimetre of a platinum surface that had been heated to its melting-point, but his suggestion has not been taken up.

The Hefner Lamp. In Germany and some other continental countries the unit of light intensity is the Hefner lamp, which burns amyl acetate and the intensity of which is exactly ·090 of the intensity of the Vernon Harcourt 10 candle-power pentane lamp. The intensity of the lamp is thus rather small for practical purposes, and the colour of the flame is redder than that of most light sources, but on account of its simplicity and cheapness and the fact that it can be bought certificated by the

Reichsanstalt the lamp has found a wide use. It is represented in fig. 263.

In using the lamp the wick is cut level with the top of the tube. Then after the lamp is lit the height of the flame is adjusted by screwing up the wick. It should extend exactly 40 mm. above the top of the tube.* The sighting arrangement on the top of the lamp to the right is for getting this height right; it consists of a horizontal tube with a horizontal diametral steel plate. To an observer looking through the sighting arrangement the bright core of the flame should appear just to touch the lower surface of this plate, and the faint luminous sheath of the flame should almost reach $\frac{1}{2}$ mm. above the upper surface of the plate.

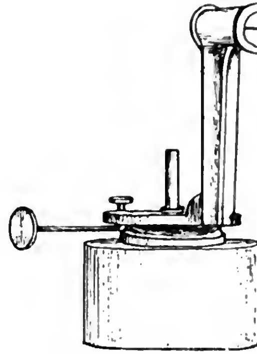


FIG. 263.

It is necessary to wait at least 10 minutes after lighting the lamp before making measurements. An error of 1 mm. in the height of the flame means an error of 3 per cent in the intensity.

The lamp is provided with a cap for the wick tube when it is not in use and a gauge for checking the difference in level between the top of the tube and the sighting arrangement.

Definition of Solid Angle. Let S be a surface which is not necessarily plane. In order to measure the solid angle subtended by S at a point P draw a sphere of unit radius with its centre at P. Let straight lines be drawn from P to every point on the boundary of S. Then these straight lines from a cone and this cone intercepts a certain area on the surface of the sphere. The solid angle subtended by S at P is numerically equal to the area intercepted by the cone on the sphere. Thus if S is a closed surface and P is inside it, the solid angle is 4π.

An Object Appears Equally Bright at All Distances. Let s be the area of a small plane surface which is at right angles to the line joining it to the eye. Let d be its distance from the eye, and let a be the area of the pupil of the eye. Let s be emitting light normally at such a rate, that if the emission were uniform in all directions the quantity of light emitted per second would be equal to L. Then the quantity of light entering the pupil of the eye per second is

$$\frac{a\text{L}}{4\pi d^2}.$$

The solid angle subtended by the object at the eye is

$$\frac{s}{d^2}$$

and the size of the retinal image is proportional to this solid angle. Hence the brightness of the object as seen by the eye is proportional to

* The National Illumination Committee of Great Britain has proposed to obtain the British Standard candle from the Hefner lamp by increasing the height of the flame; according to a determination made at Glasgow the height for this intensity is 44·37 mm.

$$\frac{a\mathsf{L}}{4\pi d^2}\Big/\frac{s}{d^2} = \frac{a\mathsf{L}}{4\pi s}$$

and is independent of d. An object consequently appears equally bright at all distances from the eye.

This is, however, only true if the object is an extended one. If it is so small that its image on the retina falls wholly on one cone,* it is a point source as far as the eye is concerned and its apparent brightness varies inversely as the square of its distance from the eye.

Star Magnitudes. Stars are point sources of this kind, and the word magnitude in connection with stars has nothing to do with their size but refers solely to their brightness. The stars visible to the naked eye were divided by Hipparchus and Ptolemy into six magnitudes, the sixth magnitude containing the faintest and the first the brightest stars. Owing to the invention of the telescope a vast number of additional stars has been rendered visible; the classification has been extended to take these in and has been refined; we say now, for example, that a star is of the magnitude 3·4. When there is a difference of unity in the magnitudes of two stars, the one emits 2·5 times as much light as the other; thus a star of the first magnitude emits $(2{\cdot}5)^5$ times or approximately 100 times as much light as a star of the sixth magnitude. There are special photometers for measuring the magnitude of stars; these depend either on visual observation or photography. In one of the visual instruments, the Zöllner photometer, the star is compared with an artificial star, the intensity of which is varied by means of a rotating nicol.

The brightness of the sky is not increased by a telescope but the brightness of the stars is. Hence by using a telescope bright stars can be made visible at daytime.

Brightness of Image Formed by a Lens. Suppose that a lens is used to form an image of the small plane surface mentioned in the second last section, and that this image lies on the straight line joining the small plane surface and the eye, and has an area of s_1. Then by the reasoning of the same section the brightness of this image is given by

$$\frac{a\mathsf{L}_1}{4\pi s_1},$$

where L_1 would be the light emitted per second by the image, if its emission were the same in all directions as it is in the direction of the eye. Let u, v, and l be the object distance, image distance, and radius of the lens. Then

$$\frac{s_1}{s} = \frac{v^2}{u^2}.$$

The lens subtends at the object a solid angle proportional to l^2/u^2, and at the image a solid angle proportional to l^2/v^2. The light falling on the lens is proportional to $(l/u)^2\mathsf{L}$, and the light falling on the image and issuing from it on the other side is proportional to $(l/v)^2\mathsf{L}_1$. Neglect the absorption and reflection losses caused by the lens and equate these two expressions. Then

$$(l/u)^2\mathsf{L} = (l/v)^2\mathsf{L}_1 \quad \text{or} \quad \mathsf{L}/u^2 = \mathsf{L}_1/v^2.$$

Thus the expression for the brightness of the image

$$\frac{a\mathsf{L}_1}{4\pi s_1} = \frac{a\mathsf{L}}{4\pi s},$$

which was the expression for the brightness of the object. This result has,

* Cf. p. 333.

however, been obtained on the assumption that no light was lost either by reflection at the surfaces of the lens or by absorption in its material. In practice the brightness of the image is consequently less than the brightness of the object. The same may be shown to be true with regard to the image formed by a mirror. Hence an image of an extended object formed by a lens or a mirror can never appear brighter than the object itself.

Spectrophotometry. So far nothing has been said about the colour of the light. If the difference in colour between two sources is slight, they can be compared with an ordinary photometer. If the difference in colour is pronounced, the comparison must be made by means of auxiliary sources of intermediate tints, the one source being compared with a second, this with another, and so on, the colour of these intermediate sources being gradually altered until finally it is not much different from the colour of the other source under comparison. Consistent results are obtained in this way, provided that the observers have normal colour vision. Sources of widely different colour can also be compared directly by means of the flicker photometer which is described on p. 348.

The most satisfactory way of comparing the intensities of two differently coloured light sources is to resolve their light into spectra and measure the relative intensities of these spectra wave-length by wave-length. The instruments used for this purpose are called spectrophotometers; a great number of different types of spectrophotometer have been made, although few of them have been much used. They differ among themselves chiefly in the means used to vary the intensities of the two spectra. This may be done by producing the two spectra from separate slits and altering the widths of these slits, or by the use of rotating sectors or wedges of neutral tinted glass, or finally by the use of polarised light. The difficulty in the way of regulating the brightness of a spectrum by altering the width of the slit is, that not only the intensity but also the purity of the spectrum is being altered, and, if the slit is opened very wide, we obtain a distinct change of colour. Also, when the slit width is small, owing to diffraction effects the intensity is not strictly proportional to the width of the slit.

In the rotating sector method an opaque disc cutting the path of the one beam of light at right angles is rotated at a high speed by means of a motor. The disc consists of sectors which can be pushed one behind the other, and consequently an open sector of variable angle formed. The disc is rotated so fast that no impression of flicker is produced, and the intensity in the field is perfectly steady and is proportional to the angle of the open sector. The disadvantages of this method are that it cannot produce very small diminutions of intensity, because there must always be a sector of appreciable angle left to shove the others in behind, also the motor is a complication.

Wedges of neutral tinted glass are always of small angle, and are cemented to wedges of equal angle made of transparent glass of equal index of refraction, so as to form a plane parallel plate. This plate is placed in front of the slit with the edges of the wedges parallel to the latter, and the intensity of the beam diminished by moving it at right angles to the slit, so as to bring a greater thickness of absorbing glass and a smaller thickness of transparent glass in front of the slit. The plate does not deviate the rays in any way; it only absorbs them. The disadvantages of this method are that it is not easy to get a perfectly neutral tinted glass, i.e. one that absorbs all the colours of the spectrum in the same degree, also that it is not an absolute method; the wedge has to be calibrated by other methods.

§ Much the most popular method of comparing the intensities of the beams has been by plane polarising them at right angles to one another and then altering their relative intensity by the rotation of a nicol. I have had considerable experience in the use of spectrophotometers and have designed and made one that employs this principle. This instrument wastes no light, gives as great accuracy as other instruments, and has in addition the advantage of great simplicity.

The only points in which it differs from an ordinary spectrometer are (1) that there is a nicol mounted in the eyepiece, which rotates together with the eyepiece about the axis of the telescope and the position of which can be read on a divided circle, and (2) that there is a prism of peculiar construction in front of the slit. This prism is

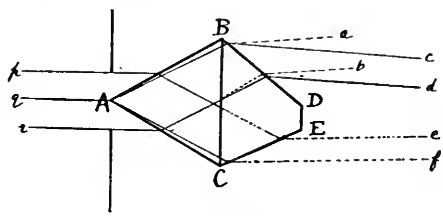


FIG. 264.

represented in fig. 264. The part ABC is made of glass for which $\mu_D = 1{\cdot}526$, the sides AB, BC, CA being each 2 cm. long. It is cemented to a prism of Iceland spar, BDEC, cut with its axis perpendicular to the plane of the paper. The angle D is 127° 12′, E is 115° 49′, and BCE is 64° 11′.

The action of the prism may be better understood by supposing the beams of light to go in the reverse direction—from the object glass of the collimator to the slit. The beam *qr* is broken into two by the Iceland spar prism, *cd* being the ordinary beam and *ab* the extraordinary. The beam *pq* is broken into two, but only the extraordinary *ef* emerges, the ordinary being totally reflected at the surface CE. The beams *ef*, *cd* meet 15 cm. out in an elliptical spot of light measuring 2·0 cm. by 2·4 cm. the long axis being vertical. The beam *ab* is quite 2 cm. clear.

If, now, we have as source of light an incandescent mantle behind a screen with an aperture at the proper place not much larger than

2·0 cm. by 2·4 cm., and if we look into the eyepiece, we see two spectra, one above the other, polarised at right angles to one another. The ordinary component of the lower beam misses the slit entirely, while the extraordinary component of the upper beam misses the object glass of the collimator. The relative intensity of the two spectra is altered by the rotation of the nicol eyepiece.

In order to measure the fraction of light transmitted through a piece of coloured glass or a glass cell containing a solution, it is placed in the path of the upper beam and the intensities of the spectra matched in two adjacent quadrants. Let the difference of the readings on the divided circle be 2α. It is then placed in the path of the lower beam and the intensities again matched in the same two quadrants. Let the difference of the readings on the divided circle be in this case 2β. Then the fraction of the light transmitted is either

$$\frac{\tan \alpha}{\tan \beta} \quad \text{or} \quad \frac{\tan \beta}{\tan \alpha},$$

whichever is smaller than unity.

To prove this suppose the piece of glass is taken away, let OA and OB be the directions of vibration in the two halves of the field, and let a and b be respectively the amplitudes of the upper and lower beams before they fall on the eyepiece nicol. a is not equal to b, because, although the light emitted by the source is not polarised in any way, yet the two beams suffer different reflection losses in traversing the instrument.

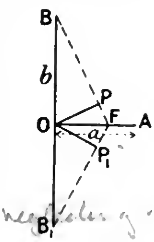


FIG. 265.

Let the piece of glass now be placed in the path of the upper beam. It diminishes the amplitude OA to OF. Let OF $= fa$. If the nicol is set to make the intensities equal, the direction of vibration after emerging from the eyepiece must be OP or OP_1, where $\angle POP_1 = 2\alpha$. OP and OP_1 are respectively perpendicular to FB and FB_1 and $OB_1 = OB$. Consequently

$$OP = b \sin \alpha = af \cos \alpha \quad \text{or} \quad \tan \alpha = \frac{af}{b} \quad . \quad . \quad (4)$$

When the piece of glass is placed in the path of the lower beam, OB and OB_1 are diminished to OG and OG_1, where $OG = OG_1 = bf$, and when the nicol eyepiece is set to match the intensities, the direction of vibration of the light after emerging from it is given by OQ and OQ_1, where $\angle QOQ_1 = 2\beta$. Consequently

$$OQ = bf \sin \beta = a \cos \beta \quad \text{or} \quad \tan \beta = \frac{a}{bf} \quad . \quad (5)$$

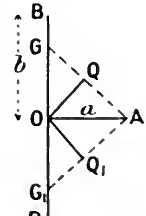


FIG. 266.

Combining (4) and (5) we obtain

$$\frac{\tan \alpha}{\tan \beta} = \frac{af}{b}\frac{bf}{a} = f^2.$$

f^2 is, of course, the fraction of the incident light transmitted by the piece of glass, i.e. the ratio of the intensity after transmission to the intensity before incidence. f is the ratio of the amplitude after transmission to the amplitude before incidence.

After f^2 has been determined for one point in the spectrum, the position of the telescope is shifted and f^2 determined in succession for other points. The results can be shown in a curve giving f^2 as a function of λ. There are screens in the focal plane of the eyepiece for pushing in from the sides so that only a narrow strip of the spectrum remains visible; these screens are very useful when f^2 varies rapidly with λ.

To compare the distribution of intensity in the spectra of two sources with this spectrophotometer one of the sources is made to illuminate a piece of ground glass measuring 2·0 cm. by 2·4 cm., which is placed 15 cm. out, while the lower beam from this piece of ground glass is stopped by a screen and replaced by a beam from a second piece of ground glass, this second piece being illuminated from the side by the second source.

A Simple Means of Spectrophotometry. Most laboratories do not possess a spectrophotometer, and the object of this section is to describe how rough spectrophotometry may be done with the simplest apparatus. All that is required is a spectroscope with a total reflection prism in front of the slit, one of these prisms which are used to observe two spectra, say the solar spectrum and the Na spectrum, at the same time; with such a prism the solar spectrum is viewed direct and occupies one-half of the field, while the rays from the sodium flame come in from the side at right angles and are totally reflected into the slit, and the sodium spectrum occupies the other half of the field.

Fig. 267 shows how this apparatus is arranged for spectrophotometry. S is the collimator slit, G_1 and G_2 two exactly similar ground glass plates, and L_1 and L_2 the two sources of light under comparison. The total reflection prism is shown in position immediately in

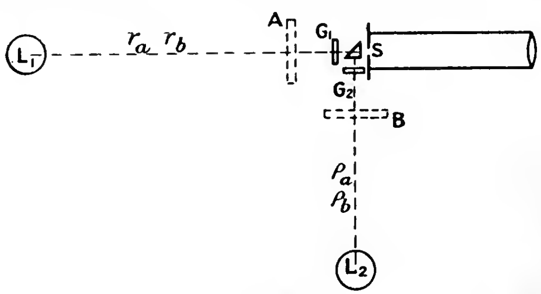


FIG. 267.

front of the slit. The rays from the ground glass G_1 pass to the slit either over or under the prism. The rays from the ground glass G_2 pass to the slit by reflection in the prism. The intensities are matched simply by the inverse square law. If for a certain colour the intensities of the two spectra are equal, then the ratio of the intensities, as far as light of this colour is concerned, is given by the ratio of the squares of the distances from their respective ground glasses, after a correction has been made for the reflection loss (8 per cent.) suffered by L_2 in the prism. This reflection loss can be checked by interchanging the positions of the sources. The distance of each source from its ground glass is measured by a metre stick not shown in the diagram.

If the fraction of light transmitted by a coloured glass plate is to be determined, the best way of proceeding is to place it in the position shown by the one dotted plate (fig. 267) and determine the ratio of the distances of the two sources, then place it in the position shown by the other dotted plate and again determine the ratio of the distances. The smaller ratio divided by the larger one gives the fraction of the light transmitted.

For let f be the ratio of the amplitude of the light transmitted by the plate to the amplitude of the light incident on it. Let I_1 and I_2 denote the intensities of the sources L_1 and L_2, and let r_a and ρ_a denote their distances when the plate is in the position A, and r_b and ρ_b their distances when the plate is in the position B. Then when the plate is in the position A

$$\frac{I_1 f^2}{r_a^2} = \frac{I_2}{\rho_a^2},$$

and when it is in the position B

$$\frac{I_1}{r_b^2} = \frac{I_2 f^2}{\rho_b^2}.$$

Dividing the first of these equations by the second we obtain

$$\frac{I_1 f^2}{r_a^2} \Big/ \frac{I_1}{r_b^2} = \frac{I_2}{\rho_a^2} \Big/ \frac{I_2 f^2}{\rho_b^2}$$

or

$$\frac{f^2 r_b^2}{r_a^2} = \frac{\rho_b^2}{f^2 \rho_a^2},$$

which gives, on extracting the square root,

$$f^2 = \frac{\rho_b r_a}{r_b \rho_a}.$$

The chief source of error in using this simple apparatus is stray light. Each ground glass must be illuminated by its own source and by nothing else, and to ensure this it is necessary to mount a number of dead-black screens. The positions of these screens will depend on the apparatus employed, the lighting of the room, etc., and must be left to the skill of the experimenter. It is possible to obtain results right to 2 per cent, but this degree of accuracy requires much skill and experience. Indeed there is hardly a better test of experimental ability than to measure the fraction of light transmitted by a piece of coloured glass throughout the spectrum by this method. The apparatus is of the simplest and the fitting up and arrangement can be done by any amateur, but, although the knowledge required does not go beyond the elementary principles of optics, still there are numerous sources of error to be explored.

Spectrophotometric Results. Suppose that an aqueous solution of salt of concentration c measured in gram molecules per litre is placed in a parallel-sided glass cell of internal thickness d, and the quantity f^2 is measured for this cell, then it includes (1) the loss due to reflection at both of the glass-air surfaces; also (2) the loss due to reflection at the glass-water surfaces; (3) loss due to possible absorption of light in the glass; and finally (4) loss due to absorption of light by the solution. If, however, when we place the glass cell with the solution in the path of the one beam, we place a similar glass cell filled with water in the path of the other beam, the first three losses are approximately the same for both cells, and consequently the quantity f^2 represents only the ratio of the intensity on the plane Q to the intensity on the plane P. The planes P and Q are inside the solution but as close to the surface as possible. It is found experimentally that in this case f^2 can be represented by

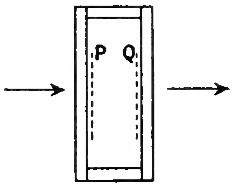


Fig. 268.

$$10^{-Acd} \quad . \quad . \quad . \quad . \quad (6)$$

where c is the concentration of the solution as already specified, d is the distance between the two planes, and A is a constant which varies with the wave-length. This fact is sometimes referred to as Lambert's law. A is called the molecular extinction coefficient of the dissolved substance.

If we increase d by a small increment δ, then the fraction transmitted is

$$10^{-Ac(d+\delta)} \quad \text{or} \quad 10^{-Acd}\,10^{-Ac\delta}.$$

The additional thickness δ transmits the same fraction $10^{-Ac\delta}$ no matter what the value of d is.

If in (6) we double c and half d, or treble c and decrease d to one-third of its value, then their product remains constant, and according to a law known as Beer's law the fraction transmitted should remain the same. This law is not always true, for in certain cases A varies with c, and consequently in these cases the product Acd does not remain constant, even if the product cd does.

The curves * represented in figs. 269 and 270 give A as a function of λ for the bromides, chlorides, nitrates, and sulphates of cobalt and nickel. The values for the visible spectrum were taken with the spectrophotometer described on p. 324 and the values for the infra-red were taken with a linear thermopile. It is impossible to go further into the infra-red with aqueous solutions because the water itself

* The curves represented in figs. 269, 270 and 271 are from a series of papers in the Proc. Roy. Soc. Edin., 1910-13, by John S. Anderson, Alex. R. Brown, Chas. Cochrane, Alex. H. Gray, and the author.

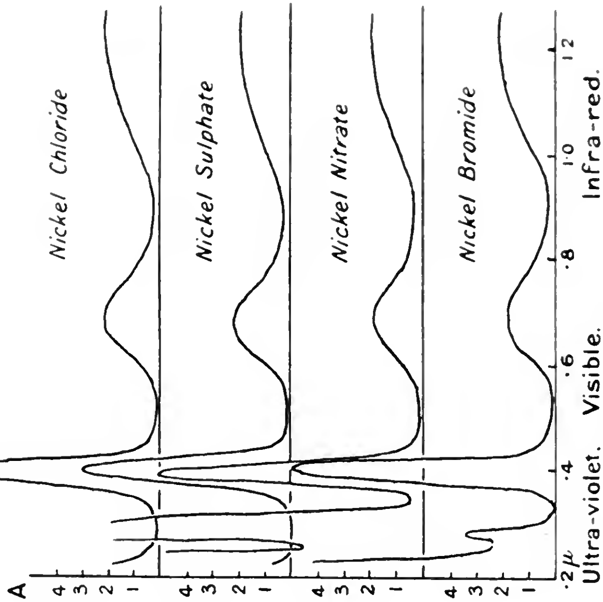


FIG. 270.

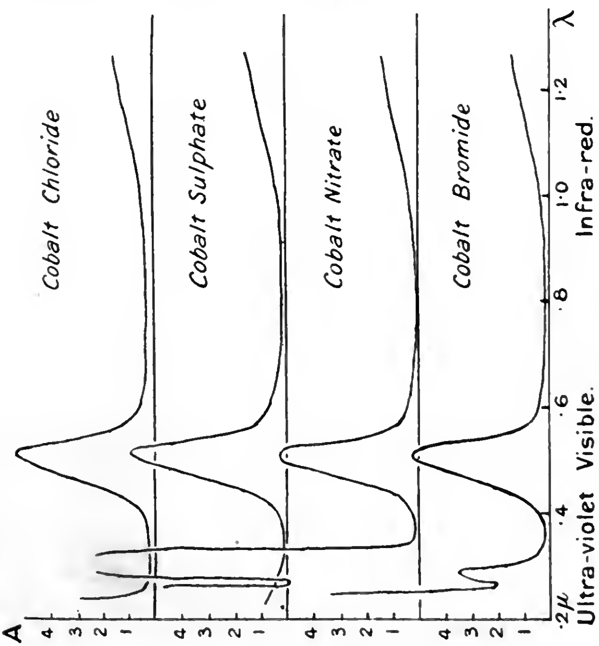


FIG. 269.

absorbs very much more in that region than the dissolved substance does. It is comparatively easy to determine f^2 with a thermopile; it is necessary only to take the ratio of the galvanometer deflections when cells filled with the solution and with water are placed in succession in front of the slit. The values for the ultra-violet were obtained with an apparatus depending on the inverse square law, in which equal intensity of light was measured by equal blackening in the same time of two neighbouring areas on a photographic plate.

It will be noticed that the values of A for the salts shown can be obtained approximately by adding the parts due to the base and to the acid radical. The cobalt salts have all the same band at ·510μ and the nickel salts have all the same bands at ·405μ, ·690μ, and 1·21μ, while the nitrates have a band at ·302μ and the bromides a band at ·285μ. Absorption spectra of salts are however not always so additive.

The nitric esters, i.e. salts of nitric acid with organic bases, do not show the band at ·302μ, and Ostwald and other physical chemists maintained vigorously that the additive property of the absorption spectra of inorganic salts could be explained, where it occurred, by the assumptions that in dilute solutions the salts were resolved into ions, and the spectrum of the solution was the superposition of the spectra of the ions. In dilute solutions the colour was due solely to the colour of the ions. This statement is still to be found in certain standard books on physical chemistry. It is impossible here to go into the evidence on the subject, but it should be stated emphatically that Ostwald's statements were based on hasty and inaccurate work. In some cases, e.g. salts of the rare earths, the spectra do not become additive even at the greatest dilutions. In cases where the spectra are additive at great dilution they are still approximately additive when the solutions are concentrated and consequently not ionised. In the case of the curves shown on p. 329 the solutions were too concentrated for the similarity to be explained by ionisation.

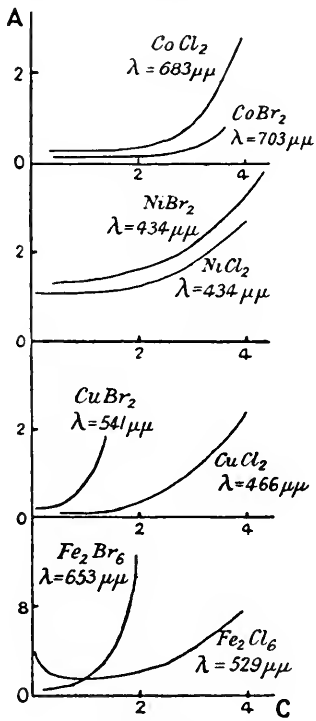


FIG. 271.

Aqueous solutions of the chlorides and bromides of cobalt and copper change colour when they are diluted, i.e. the value of A changes with dilution, and this change in A has been attributed to the molecules ionising. There is, however, no doubt that it is due to dehydration of the molecules as the solution becomes more concentrated. Similar changes occur in the case of ferric chloride and bromide and nickel chloride and bromide, although this is not generally known, and in fig. 271 all these changes are represented in each case for the wave-length for which they are most pronounced. The abscissæ give concentrations in gram molecules per litre, and the ordinates represent A. The last point on each curve was for a saturated or nearly saturated solution. In the case of ferric chloride the absorption increases for dilute solutions owing to the formation of colloid hydroxide, but in every other case it is constant for dilute solutions, showing that the change has nothing to do with ionisation. For example, in the case of cobalt chloride the molecular conductivity for $c = 2$ is less than half the value for infinite dilution, but for both these concentrations the values of A are practically the same.

The theory of electrolytic dissociation has been of no use in the interpretation of absorption spectra, although of course it has been of great value in other fields.

Another source of error that has retarded knowledge almost as much as the theory of the colour of the ions has been the statement known as Kundt's law. Kundt in 1874 dissolved substances that produced absorption bands, using different solvents, and found that the position of the band varied with the solvent. He came to the conclusion that the greater the refraction and dispersion of the solvent, the further was the band displaced towards the red. This statement is not true ; the band moves as often the one way as the other, and generally there is a marked change in intensity and shape, in comparison with which the shift in position can be neglected.

EXAMPLES.

(1) The distance between two incandescent lamps of 8 and 16 candle-power is 6 feet. Show that there are two positions on the line joining the lamps in which a screen may be placed so as to receive equal illumination from each of them, and determine these positions.

(2) If the light of the full moon is found to produce the same degree of illumination as a standard candle does at a distance of 4 feet, what is the equivalent in candle-power of the moon's light? The moon is distant 60 times the earth's radius.

(3) Two stars of the fourth magnitude are so close together that they appear as one. What is the magnitude of the two stars combined?

(4) If the naked eye can just see a sixth magnitude star and the diameter of the pupil of the eye is 3 mm., what is the magnitude of a star which is just visible with the Yerkes' telescope, the diameter of the object glass of which is 40 inches?

(5) The star Algol undergoes a periodical partial eclipse, during which its intensity sinks to $\frac{1}{6}$ of its usual value. The eclipse lasts altogether about $8\frac{1}{2}$ hours from start to finish, and the time that elapses between two successive

eclipses is 2 days 21 hours. The phenomenon has been explained as due to a small dark star rotating round a bright one (two bright ones rotating round one another would give two minima). Calculate the ratio of the radius of the dark star to the radius of the light one. Assume that they are both spherical, and that the disc of the bright one is uniformly bright. As the intensity remains constant for about 20 minutes at the minimum, it may be assumed that the one disc falls wholly inside the other one at the minimum.

CHAPTER XIX.

THE EYE AND COLOUR VISION.

The Eye. The eye consists practically of a spherical chamber with a circular opening in the front; by means of a system of lenses the light entering this opening forms an image on the back of the chamber just as in the case of a photographic camera.

Fig. 272 represents a section of the human eye. It is surrounded by a coating S called the sclerotic. The front portion of this coating, C, is transparent and is called the cornea. L is the crystalline lens, which is attached to the walls of the eye by the ciliary muscle. In front of the lens is a diaphragm, I, called the iris, which is coloured; it is the colour of the iris that is meant when the colour of an eye is referred to. In the centre of this diaphragm there is a circular aperture called the pupil. The space A between the lens and cornea, i.e. the anterior chamber, is filled with a watery liquid containing a little salt in solution, which is called the aqueous humour. The space V behind the lens, i.e. the posterior chamber, is filled with a transparent gelatinous substance termed the vitreous humour.

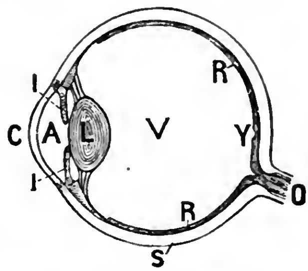


FIG. 272 (from Watson's "Physics").

After passing through the cornea, aqueous humour, lens and vitreous humour the rays fall on the retina R. The retina consists principally of a network of nerve fibres connected with the brain by the optic nerve O. In the retina there are two kinds of vision cells termed rods and cones. In the retina directly behind the pupil is the yellow spot Y, or macula lutea, which has a depression in the centre called the fovea centralis; here vision is most distinct. There are no rods in the fovea centralis. The point in the retina at which the optic nerve enters is not sensitive to light and is called the blind spot.

Usually when the eye is at rest it is adjusted so that the image of a distant object is in focus on the retina. If then it is turned to a nearer object, its pose must be altered, it must be accommodated to the nearer object in order that the latter may be clearly seen. This is done by a forward motion of the lens and an increase in curvature of both its surfaces. The range of accommodation of the eye is limited. The normal eye sees small objects best when they are at a distance of 25 or 30 cms. This distance is called the distance of distinct vision because, if the object is nearer, an exertion is required to focus it, and if it is further away, although quite as sharp, it is smaller and the detail cannot be seen so well. Helmholtz found that when the focus of the eye was changed from infinity to the distance of distinct vision, the radius of curvature of the front surface of the lens changed from 10 mm. to 6 mm., and the radius of curvature of the back surface from 6 mm. to 5·5 mm.

The crystalline lens is biconvex and colourless, and consists of many layers of different density. The outmost layer is soft, the inmost one harder. The index of refraction of the outmost layer is 1·405, of the middle layer about 1·429, and of the inmost layer 1·454. The indices of refraction of the aqueous and vitreous humours have about the same value, 1·34. When the cornea is adjusted for distant vision its radius of curvature is about 7·8 mms.

The index of refraction of the cornea is nearly the same as that of the aqueous humour. Consequently rays entering the eye suffer refraction mainly at three surfaces, the outer surface of the cornea and the two surfaces of the lens. There is in addition to this the continuous refraction experienced in passing from layer to layer of the lens. The eye thus consists of a system of coaxal spherical refracting surfaces and the theory of Chapter III can be applied to it. As the initial medium is air and the final medium vitreous humour, the nodal points do not coincide with the principal points. The positions of the cardinal points for the average eye focussed for parallel light are, according to Listing, as follows: The first principal point is 1·7 mm. behind the front surface of the cornea, and the second principal point is 2·1 mm. behind the same surface. They are thus both situated in the anterior chamber of the eye. The first principal focus is 13·7 mm. in front of the cornea, and the second principal focus is on the retina 22·8 mm. behind the cornea. The first nodal point is situated in the lens ·2 mm. in front of its back surface, and the second nodal point is situated in the posterior chamber ·12 mm. behind the back surface of the lens. The optical axis of the system does not pass through the centre of the yellow spot.

The ophthalmometer is an instrument used for measuring the curvature of the cornea. The principle it employs is that described on p. 71 for obtaining the radius of curvature of a convex mirror. The formula in that case was

$$r = \frac{2l\mathrm{D}}{\mathrm{L} - 2l},$$

where L was the length of an object placed in front of the mirror, D was the distance of the object from the mirror, and l the length of the image formed in the mirror. In measuring the curvature of the cornea

l is always much smaller than L, and hence can be neglected in comparison with the latter. The formula therefore becomes simply

$$r = \frac{2l\text{D}}{\text{L}}.$$

The image in the eye is regarded through a double-image prism and is thus seen double. The length L is altered until the two images just touch; l is then known from the construction of the instrument and D can be measured.

For an object to be seen clearly its image must fall on the middle of the yellow spot. Then two points can be distinguished when they subtend an angle of 1′ at the eye, that is, when their images are ·005 mm. apart on the retina. Rays falling on the peripheral regions of the retina form blurred images, and the objects emitting them cannot be seen distinctly.

The image formed on the retina is, of course, an inverted one, but this causes no confusion since the impression it produces is always associated mentally with the upright position of things.

The normal eye suffers both from spherical aberration and chromatic aberration. The presence of the former can easily be shown. If a piece of cardboard is pierced with a pinhole and this pinhole held close up to the eye and a page of printing looked at through it, it is found that the printing can be read easily when closer to the eye than the near point. The focal length is thus less for the axial rays than for the marginal rays, the opposite to what happens in the case of the single convex lens, or, in other words, the eye is over-corrected for spherical aberration.

When the eye is adjusted for parallel light the focus of the violet rays lies 0·43 mm. nearer the lens than the focus of the red rays does. This difference of focus can be shown very well with certain kinds of cobalt glass which let through only the blue and violet and the extreme red end of the spectrum. If we look through such a piece of glass at the filament of an electric glow lamp we see two images superimposed, a red one and a blue-violet one. If the filament is at a great distance, the eye focusses involuntarily on the red image which consequently appears surrounded by a blue-violet haze. If the filament is so near that the eye cannot focus on the red, the blue-violet image is seen surrounded by a red haze.

Defects of Vision. There are four defects of the eye of frequent occurrence, which are remedied by the use of spectacles. These are (1) Myopia, or Short Sight; (2) Hypermetropia, or Long Sight; (3) Presbyopia; (4) Astigmatism.

In myopia, or short sight, the focal length of the eye is too small and parallel rays are brought to a focus in front of the retina. Consequently distant objects cannot be seen distinctly. Near objects can be seen quite well, and the distance of distinct vision is less than in the case of the normal or emmetropic eye. The defect is corrected by the use of a concave spectacle lens.

In hypermetropia, or long sight, the focal length of the eye is too great, and, when the eye is relaxed, parallel rays are brought to a focus behind the retina. By an act of accommodation the focus can be brought on to the retina in the case of distant objects and consequently they can be seen clearly. But, even when the lens is curved as much as possible, the images of near objects still fall behind the retina and appear blurred. In the hypermetropic eye the distance of distinct vision is greater than for the normal eye. Hypermetropia is corrected by the use of a convex spectacle lens.

If d is the distance of distinct vision for the hypermetropic eye, and D for the normal eye, then the focal length of the correcting lens is given by

$$\frac{1}{f} = \frac{1}{d} - \frac{1}{D},$$

for its function is to change a real object situated at a distance D to a virtual object at a distance d on the same side of the lens, and it is this virtual object that is viewed by the eye. The same formula gives the focal length of the spectacle lens to be used with a myopic eye; in that case the sign comes out different.

Presbyopia is a loss of accommodating power occurring in the case of elderly people owing to the stiffening of the muscle that alters the curvature of the crystalline lens. People afflicted with it can see distant objects clearly but not near objects. Sometimes the accommodation almost entirely disappears. Convex spectacle lenses, usually powerful ones, must be used for reading and writing with and less powerful ones for getting about with.

In astigmatism the focal length of the eye is different in two planes at right angles to one another, that is, if on a card two sets of lines are ruled parallel and at right angles to one another and the card is rotated so that the directions of the lines correspond with the above two planes, only the one set of lines can be seen distinctly at once. If the other set is to be seen distinctly, the accommodation of the eye must be altered and the first set put out of focus. Astigmatism is due to the surfaces of the lens and the cornea, principally the surface of the cornea, not being symmetrical about their axis. It is corrected by the use of cylindrical, sphero-cylindrical, and toric lenses. A sphero-cylindrical lens is one, one surface of which is spherical and the other surface of which is cylindrical. A toric or toroidal or more strictly a plano-toric lens has one surface plane and the other with different curvatures in its two principal meridians.

The power of spectacle lenses is always measured in diopters. The power of a lens in diopters is obtained by dividing 40 by its focal length in inches or, more accurately, by dividing 100 by its focal length in cms. A + is prefixed to the result if the lens is convex, and a − if it is concave. Spectacle lenses are mostly biconvex or biconcave and made of glass of index of refraction 1·52, hence the focal length can be determined from the radius of curvature of one side. There is an instrument for this purpose, the Geneva lens tester, something like the spherometer, which rests on the surface of the lens and in which a central point is pressed down to make contact. When contact is made, the power of the lens can be read off on a scale directly in diopters.

§ When the ciliary muscle is quite relaxed, the eye is focussed on its far point or *punctum remotum*. In normal vision the far point is at infinity. When the ciliary muscle has contracted as much as it can, the eye is focussed on its near point or *punctum proximum*. The near point is, of course, closer to the eye than the distance of distinct vision. The near point can be determined by noting the shortest distance at which very small type is legible. The distance between the near point and far point of an eye is called its range of accommodation.

When the eye is adapted for the far point, the change in power necessary to accommodate it to the near point is called its amplitude of accommodation. Or, in other words, the amplitude of accommodation is the power of a lens which enables a perfectly relaxed eye to see an object situated at its near point. With the advance of age the elasticity of the eye lens and the amplitude of accommodation diminish, as is shown by the following table :—

Age.	*Average Value of the Amplitude of Accommodation in Diopters.*
10 years	14
15 ,,	12
20 ,,	10
30 ,,	7
40 ,	4·5
50 ,,	2·5
60 ,,	1·0
75 ,,	0·0

Binocular Vision. When we look at a near object, the lines of vision of the two eyes have to converge towards it and an act of muscular accommodation has to be made. The magnitude of the slight exertion involved in this act helps to give us an idea of the distance of the object. But our estimation of distance is principally made in another way altogether. The image formed on the fovea of the one eye is not exactly the same as the image formed on the fovea of the other, because the object is regarded by the two eyes from different standpoints. The difference is, of course, easily noticed when we look through a window and close first one eye and then the other ; the window frame appears to move relatively to the background. This difference in the two images we associate with solidity ; if the two images are exactly the same, as is the case when we look at a picture, the images on the retinas of the two eyes fuse together perfectly, and give a sensation of flatness, and the objects in the picture do not stand out in relief.

The Stereoscope and the Stereo-Micrometer. The stereoscope is an instrument used for looking at photographs with, which gives them a wonderful sense of relief and solidity. The photographs must be taken in pairs, each pair of the same object but from slightly different standpoints. Fig. 273 illustrates the principle of the instrument. *ab* and *a'b'* are the two photographs. *ab* is looked at by the one eye through the half lens M, and *a'b'* by the other eye through the half lens N, which is of the same power as M. Each of these lenses produces a virtual enlarged image but acts also as a prism and deviates the rays. The two images thus appear at AB on the top of one another and, as owing to the photographs being taken from different points the images do not superimpose perfectly, the observer sees one picture which seems to stand out in relief.

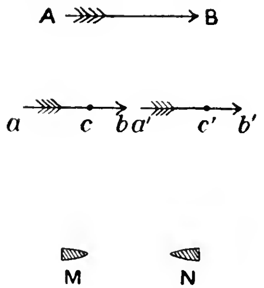


FIG. 273.

Photographs for the stereoscope might be taken with an ordinary camera by placing it first in the one position and then in the other, provided of course that the object did not move. They are, however, usually taken simultaneously by means of special cameras with two equal lenses mounted a short distance from each other side by side.

The stereoscope has developed recently into an important scientific instrument, the stereo-micrometer. If two similar pointers are placed at *c* and *c'* (fig. 273) in the plane of the photographs so that their images superimpose, they will appear to the observer as one. If the one pointer is displaced to the side, the degree of superposition of the two images will alter, and consequently the distance of the pointer from the observer will appear to alter. If the displacements of the one pointer to the side are read on a scale, it can be used to measure the distances from the observer of different objects in the photograph, for the image of the pointer can in succession be made to appear as far away as these different objects, and the scale calibrated with a photograph the distances of the objects in which are known.

This is the principle of the stereo-micrometer, which has been used amongst other purposes for determining the heights of the mountains of the moon and the depths of their craters. Two photographs of the surface of the moon were taken from different angles * and inserted in the instrument. They gave the sensation of relief, and, by bringing the mark first to the same distance as the bottom of the crater and then to the same distance as the top of its walls and reading the scale both times, the depth of the crater was obtained.

Stereoscopy has also been used for the detection of planetoids. If two photographs are taken of the same portion of the heavens on successive nights and are compared, the planetoids or small planets can be detected by the fact that they have moved relatively to the stars,

* While on the whole the moon keeps the same face towards the earth, yet it oscillates slightly about this stationary position. Hence it is seen from different angles at the same point on the earth's surface on successive nights.

which all remain fixed with regard to one another. The old way of detecting them was therefore to measure the position of every object on the plate in turn—an extremely tedious process. But if the two photographs are placed in the stereoscope, the planetoids at once stand out in relief from the background of stars, and their detection is easy.

Optical Illusions. Figs. 274, 275, 276, 277, and 278 represent some optical illusions. In fig. 274 the white square upon the black ground

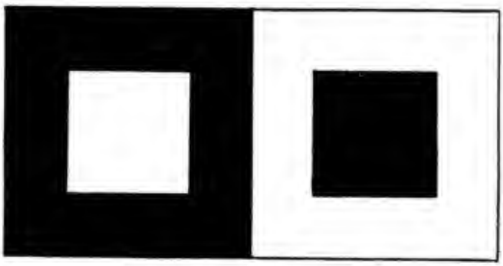


FIG. 274.

looks larger than the black square upon the white ground, although, as may be tested by actual measurement, it is really slightly smaller.

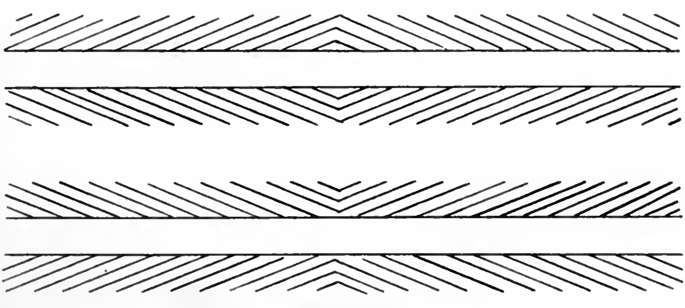


FIG. 275.

This is due to what is called irradiation. Owing to the imperfections of the eye the image of a luminous point on the retina is a circle. Hence the images of luminous points on the edges of the white square invade the black border, and similarly the white border appears to extend into the black square.

In fig. 275 the horizontal lines are parallel although they appear in the first case closer together at the middle and in the second case farther apart at the middle. In fig. 276 the vertical lines are parallel, although they appear slanted alternately in different directions. The divergence appears greater if the diagram is turned through 45°. The

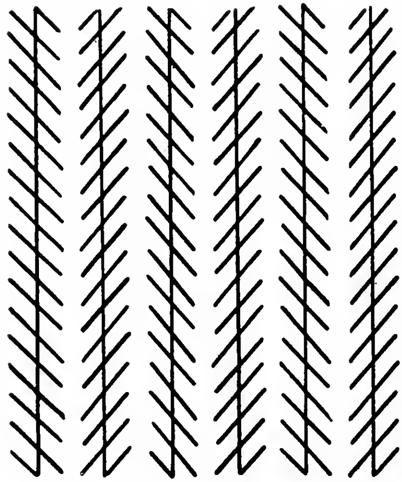


Fig. 276.

circle in fig. 277 is accurately drawn although it appears extended towards the right and compressed towards the left. The square in

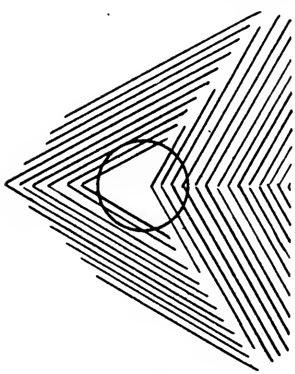


Fig. 277.

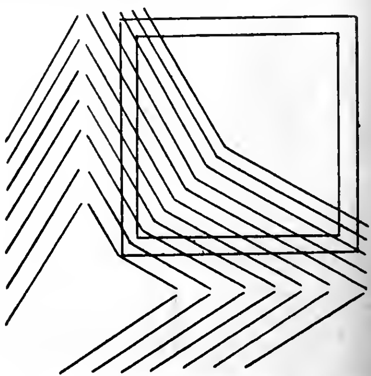


Fig. 278.

fig. 278 is a perfect square although the bottom left-hand corner appears acute.

Abney's Colour Patch Apparatus. We shall next discuss the phenomena of colour vision, and as a preliminary shall describe Abney's colour patch apparatus, with which they can be shown in a very clear way.

A plan of the apparatus is shown in fig. 279. It is somewhat similar to an ordinary spectroscope. As source of light an arc lamp is used at E. An image of the crater is focussed by the lens L_1 on the

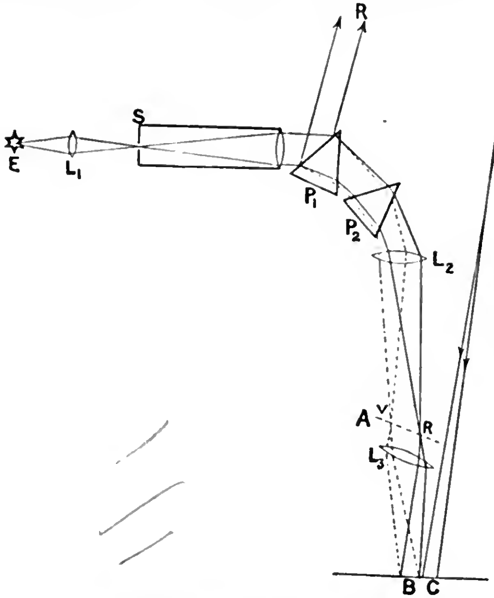


FIG. 279 (from Watson's "Physics").

slit of the collimator at S. After emerging from the object glass of the collimator the light passes through two prisms P_1 and P_2, is received by a lens L_2, and formed into a spectrum on a screen VR. V gives the violet end and R the red end of this spectrum. Suppose the screen VR removed and a lens L_3 inserted so as to form a sharp image of the second face of the second prism on the screen BC. Then, since all the colours of the spectrum fall on L_3, the colour of this image or patch, as it is called, will be white.

If in the position VR we place a screen with a slit in it and move this slit along the spectrum, then the position and shape of the patch at BC remain unaltered, but its colour varies according to the part of the spectrum in which the slit is situated. Its intensity varies according to the width of the slit.

The rays R reflected from the first surface of the first prism are reflected by a mirror, and focussed by a lens so as to fall on the screen at C, and form a white patch there side by side with the coloured patch. Since both these patches come from the same source, if the intensity of the latter fluctuates a little, the relative intensity of the two patches remains unaltered. A rotating sector is placed at D so as to alter the intensity of the white patch.

By moving a slit along the spectrum and comparing in succession the intensities of the coloured patches thus produced with the intensity of the white patch, Abney was able to measure the luminosity of the different parts of the spectrum. His results are shown in fig. 280.

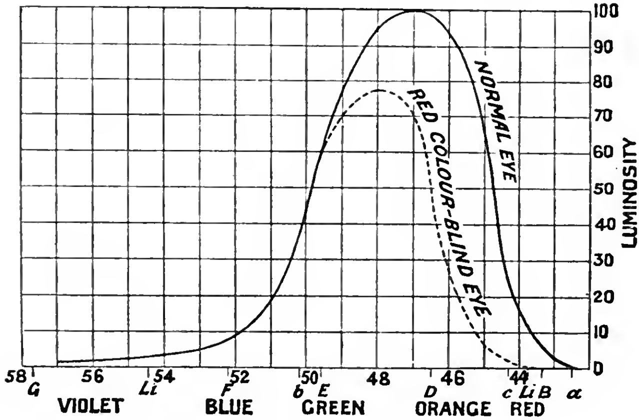


FIG. 280 (from Watson's "Physics").

The abscissæ give the positions of the slit on an arbitrary scale and the ordinates the luminosity. It will be seen that the spectrum is brightest in the yellow, and that the luminosity of the blue and violet is very small. The results are, of course, for a normal eye. For a red colour blind eye they would be represented by the dotted curve.

Colour Mixing. Suppose now that we have two slits, one in the green and one in the red. These give us two patches at BC, a green and a red one superimposed, and these two combine to form a yellow patch. This result seems at first sight very striking, because when green and red pigments are mixed the result is not yellow but a dirty brown. Also, if blue and yellow pigments are mixed, the result is green, whereas, if slits are cut in the screen so as to let a green and a yellow patch superimpose, the result is either a yellowish-white or a bluish-white. When the widths of the two slits are properly adjusted, a pure white results.

A little consideration shows that the two cases are different. When white light falls on a yellow pigment, it goes some way into it and is then diffusely reflected out, and during its passage inside it some of the constituent components of white light are absorbed. The light that comes out is coloured yellow, but this yellow is not spectrally pure ; it contains also red and green. In the same way, when white light falls on a blue pigment, it enters some distance before it is reflected out and the reflected light contains green as well as blue. If the two pigments are mixed, the incident white light is absorbed by both before it emerges again, and consequently the only constituent that emerges is that which is absorbed by neither, that is, green. Mixtures of two pigments give only the colour that is absorbed by neither, not the sum of their colours.

If, however, instead of mixing the pigments we take a disc, divide its surface into two sectors, paint one sector red and the other green, and rotate the disc at a high speed, the two colours merge owing to the rotation, and, if the angles of the sectors have been chosen properly, the disc appears yellow. That is, the rotating disc gives the same result as the colour patch apparatus. It is, however, not so satisfactory a method of mixing colours, because it is difficult to get pigments that are spectrally pure.

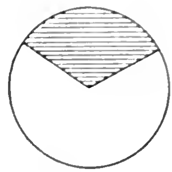


Fig. 281.

Complementary Colours. If part of the spectrum at VR (fig. 279) is blocked out with a screen, and then this part allowed to pass and the rest blocked out, the two colours which form in succession on the patch BC would obviously give white if superimposed on one another. They are said to be complementary colours, and they are composite, each containing a great range of wave-lengths.

But two monochromatic colours can also be complementary. It can be shown by having a screen at VR with two slits in it and allowing two monochromatic colours to fall on the patch, that in many cases these also give white when combined. Thus, according to Helmholtz, the following colours are complementary:—

	Wave-length		Wave length
Red	6562 A.U.	Greenish-blue	4921 A.U.
Orange	6072 ,,	Blue	4397 ,,
Yellow	5853 ,,	Blue	4854 ,,
Yellow	5739 ,,	Blue	4821 ,,
Yellow	5671 ,,	Dark blue	4645 ,,
Yellow	5644 ,,	Dark blue	4618 ,,
Greenish-yellow	5636 ,,	Violet	4330 ,,

The complementary colour to green is purple, which of course is not monochromatic, containing, as it does, both red and blue. Two colours in the spectrum, which lie closer together than complementary colours, give, when combined, one of the colours that lie between them, and this resulting colour is more saturated, i.e. contains less white in it, the closer the original colours are together.

The Young-Helmholtz Theory of Colour Vision. The table of complementary colours above shows that each different wave-length does not produce an independent sensation, otherwise the different pairs of wave-lengths would not give the same sensation of white. In this respect light waves are quite unlike sound waves. When two musical notes are superimposed, they each produce their own sensation, and the ear can detect the presence of both. But light waves seem only to affect the same common sensations though to different extents.

According to the theory of colour vision prevalent at present, the three-colour theory first given by Young and afterwards revived by Helmholtz, there are three kinds of nerve fibres in the eye. The excitation of the first of these produces the sensation of red, that of the second green, and that of the third violet. Monochromatic light falling on the retina excites usually all three sets of nerves but to different extents. If all three sensations are excited to the same extent a sensation of white is produced. Fig. 282, which was calculated by Koenig

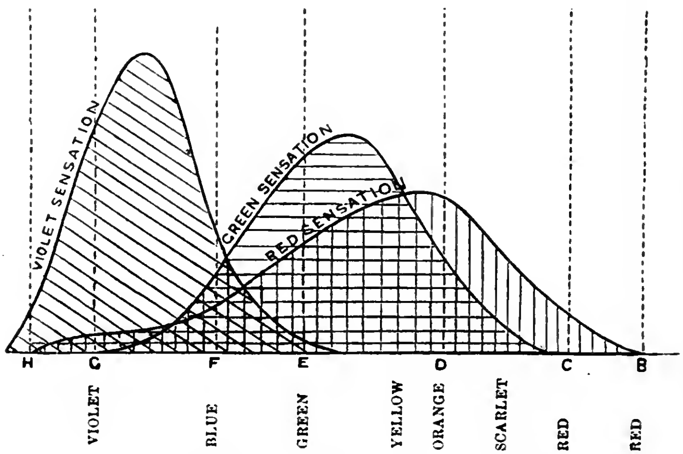


FIG. 282 (from Abney's "Colour Vision").

from colour mixture experiments, illustrates this. If light of the red hydrogen line C falls on the retina it excites only the red sensation. The sodium line excites both red and green, but the red to a greater extent. The blue excites all three sensations. If now we take any two complementary colours from the table on p. 343 and adjust their intensities until they give white, then it will be found from fig. 282 that together they always produce the three sensations to the same extent. Of course, it will be noticed that the primary green sensation cannot be stimulated without stimulating at least one of the other two.

The fact that red, green, and violet are the primary colour sensations can also be shown with a rotating disc, for by painting three sectors with these colours and varying the angles of the sectors all the natural colours can be represented.

Artists are accustomed to regard red, blue, and yellow as primary colours, but, as has been mentioned above, when pigments are mixed we do not get the sum of their colours but only the colour that is absorbed by none of them. If pigments were spectrally pure, when mixed they would always give black. Of course, not all the natural colours can be reproduced by mixing red, blue, and yellow pigments.

If we look at a bright red light for some time, until the nerves affected by red become fatigued, and then fix the eye on a white surface, the red fibres being fatigued are not affected by the red rays contained in the white, while the green and violet fibres are excited in the usual way. Hence we obtain a complementary image of the red light.

A small white object on a coloured ground appears to have the colour complementary to the ground. Thus white on a blue ground appears pink. The effect is heightened by putting a thin sheet of tissue paper over the object, but disappears if the object is surrounded by a black border. Also two complementary colours placed side by side appear heightened in intensity. According to Helmholtz, in these cases the effects are psychological, not physiological; it is our judgment not our sensation that is at fault.

Although the Young-Helmholtz theory is the prevalent one at present, it is not regarded as perfectly satisfactory, but only as the best we have got. Another theory formerly advocated was that of Hering, according to which there are three molecular processes occurring in the retina, each of which has an anabolic or constructive direction and a katabolic or destructive direction. The pair of colours black and white is associated with the one process, the pair yellow and blue with another process, and the pair red and green with the third process, black, blue, and green being associated with the constructive change, and white, yellow, and red with the destructive.

Colour Blindness. The Young-Helmholtz theory is supported by the phenomena of colour blindness. A colour-blind person is one in whom one or, in very rare cases, two of the sets of nerves are insensitive. The most common defect is insensibility to the red, or Daltonism, as it is called, after the famous chemist Dalton, who was afflicted with it; blindness to green and violet is rare. About four out of every hundred men suffer from colour blindness in some form, but only about four in every thousand women. To a red-blind person red objects appear black. Colour blindness is usually hereditary.

We can see how objects appear to a red-blind person by looking at them in the light of a mercury vapour lamp or through an aqueous solution of copper sulphate. The mercury vapour lamp emits practically no red rays, so its light cannot affect the red sensation, and copper sulphate absorbs the rays that affect the red sensation, so that the colour matches which are made under these conditions are the same as those made by a person lacking the red sensation.

Colour-blind people name the colours wrong, but the defect cannot safely be detected in this way. The usual method is by the Holmgreen wool test, introduced by Holmgreen, a Swedish physician. This consists in giving the man under examination a great number of differently coloured skeins of wool and asking him to match certain colours. The advantage of having the colour on wool is that it has then no sheen. The colour-blind man makes wrong matches. If he is shown a spectrum and is red blind, he sees it too short at the red end. If he is violet blind, he sees it too short at the violet end. If he is green blind, he sees a white part in the middle.

Many people have colour vision differing from the normal without being colour blind. The part in the spectrum where yellow changes to green often reveals this fact. What the normal call greenish-yellow, they call yellowish-green. This shows a slight deficiency in the red sensation.

§ At present the subject of colour vision is in an unsatisfactory state. There is no anatomical evidence for the existence of three kinds of nerve fibres in the eye, and there is no doubt that the popularity of the three colour theory is partly due to a predisposition to believe in primary colours, owing to the experiments in mixing pigments which we are all familiar with from childhood.

The Holmgreen wool test has been strongly attacked, and, though the Board of Trade use it, the Admiralty employ instead the Edridge-Green lantern and spectrometer tests. The spectrometer test consists in finding the number of monochromatic patches into which the candidate divides a spectrum. Dr. Edridge-Green has put forward a theory of colour vision according to which colour blindness is due to the colour perceiving centre in the brain not being sufficiently developed, but his theory suffers from the want of a clear mental picture of this colour perceiving centre.

Colour Photography. The Lippmann process of colour photography has already been mentioned on p. 151, and it has been stated that on account of its difficulty it has found little application. A picture of the spectrum taken by the Lippmann process reproduces every colour in the exact wave-length in which it was taken. Several processes of colour photography have been founded on the three-colour theory of vision and are in wide use. They do not seek to reproduce every spectral colour in light of its own wave-length, but only as a combination of the three fundamental colours, which would produce the same colour sensation in the eye. They all employ three colour filters or screens, a red one, a green one, and a blue one.

In one process three separate negatives are taken of the object to be photographed, one through each screen. A red object appears only on the negative taken through the red screen, and a green object only on the negative taken through the green screen. Yellow is partly absorbed and partly transmitted by both the red and green screens, so a yellow object comes out on the negatives taken through each of these screens. The colouring matter of the filters is selected so that the ratio of the effects on the two negatives is as nearly as possible the same as the ratio of the effects on the red and green nerves, when the colour falls directly on the eye. It is impossible to get colour filters which are theoretically accurate, and the filters in use are only a compromise; this is why blue is used for the third filter instead of the theoretically more accurate violet.

Three separate negatives are obtained, then, corresponding to the three fundamental colour sensations. From each of these negatives a positive is produced, and by combining each positive with the filter through which the corresponding negative was taken and projecting them simultaneously on a screen, so that they superimpose exactly, a picture is obtained on the screen of the original object in its natural colours. This requires what is known as a triple lantern, one with three condensers and three projecting lenses.

The Kinemacolor process of cinematography uses this principle. When the object is cinematographed, two colour filters, a red and a bluish-green one, pass in succession in front of the lens so that the exposures are made alternately through the red and green one. The negative film is then developed and the positive film made. When the positive film is projected, a red and green filter pass in succession through the beam of light, so that each positive is projected through the filter through which the corresponding negative was made. We thus have on the screen red and green pictures in such rapid succession that the colours fuse together, and all the various combination tints of the original object are produced. The method is a very ingenious one. A more accurate colour rendering would be obtained if three colour filters were used instead of two, but it is not yet possible to take and project the pictures fast enough to permit of this. Even as it is, when the rapidly moving spokes of a wheel are shown, or the legs of a galloping horse, the successive pictures are appreciably displaced so that the spokes and legs appear fringed alternately with red and green, which is very trying to the eyes.

When the three pictures in the fundamental colours are superimposed by the triple lantern, the picture is produced by what is known as the additive process. In commercial three-colour half-tone work what is known as the subtractive process is used. In this process the negatives are taken in the same way as in the additive process through the red, green, and blue filters, but they are printed directly on paper on the top of one another, each in the colour complementary to that of the filter through which it is taken. For example, the red parts of the object come out black in the red filter negative, and all the light parts of the negative are printed greenish-blue. The green parts of the object come out black in the green filter negative, and all the light parts of the negative are printed in magenta. Where the magenta and greenish-blue pigments superimpose, we obtain the colour that is absorbed by neither pigment, i.e. blue. Thus the parts of the object which come out light in the red filter and green filter negatives come out blue in the print, which is as it should be. The difficulty of the subtractive process lies in the selection of the inks.

In using the additive process it is possible to take the three separate negatives on the same plate. This method, which was first used by Prof. Joly, has been put on the market in two different modifications, the Lumière Autochrome plate and the Paget colour plate. The Lumière plate is an ordinary colour sensitive plate covered with a layer of flattened starch grains. These grains are coloured red, green, and blue, and are well mixed over the whole surface of the plate. The negative is taken through the grains, which act the part of filters. After exposure and development the image in the emulsion is reversed so that it is converted into a positive. The object can then be seen in its natural colours on holding the plate up to the light. The grains can be seen on examining the plate with a microscope.

In the Paget process the colour screen consists of a glass plate on which red, green, and blue rectangles are arranged according to a geometrical pattern. The rectangles are too small to be seen by the eye, so that when the screen is held up to the light it appears of a neutral colour. The rectangles are much larger than the grains on the Lumière plate and can be seen with a magnifying glass. The screen is placed with its colour side in contact with the emulsion of an ordinary colour sensitive plate and an exposure made. The screen is then removed, the negative developed, and a glass positive made. The positive appears like an ordinary positive except for the fact that, when examined with a magnifying glass, there is a geometrical pattern superimposed on the picture. But when the positive is placed in contact with the colour screen, so that each rectangle on it is in front of the colour through which it was taken, then the picture appears in its natural colours.

The Purkinje Effect. The maximum of the luminosity curve shown in fig. 280 is in the yellow near the D lines. If, however, as we look at a spectrum, the width of the slit is gradually decreased, the maximum of brightness gradually shifts from the yellow, and, when the spectrum has become very faint, it has reached the green. The distribution of the energy in the spectrum remains always the same, but when the light is faint the eye becomes more sensitive to green and blue than it is to yellow. This phenomenon is known as the Purkinje effect. The change takes place within the range of luminosity 150 to ·03 metre-candles. The luminosity of the spectrum is one metre-candle when the radiation from the surface which receives the spectrum is as intense as the radiation from a candle at one metre distance. If the luminosity of the spectrum is increased above the one limit or below the other, there is no further shift in the point of maximum brightness.

According to Von Kries, the Purkinje effect can be explained by assuming that the rods in the retina are chiefly responsible for vision at low intensities and the cones for vision at high intensities. The cones are sensitive to colour and have a maximum of sensitiveness in the yellow; the rods are not sensitive to colour but have a maximum of sensitiveness at a wave-length corresponding to the green. The rods are supposed to be a survival of a more elementary form of visual apparatus. The Purkinje effect explains entirely the greenish-blue appearance of objects in the moonlight; moonlight appears cold merely because it is faint.

The Purkinje effect can be shown by connecting up an incandescent electric lamp with a rheostat and gradually increasing the resistance until the light goes out. The lamp should be covered with something so as to give a large diffuse light—a piece of white cloth placed over the top of an upturned beaker for example, if the lamp happens to be lying on the table. The experiment requires a very dark room. As the current is decreased the light becomes redder and fainter, but just before it goes out it turns grey.

The Flicker Photometer. S_1 and S_2 are two light sources. AB is

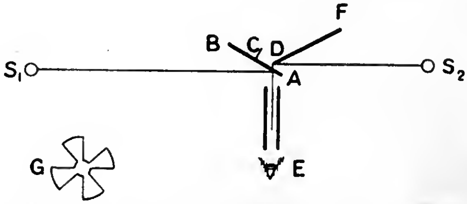


FIG. 283.

a screen in the form of a Maltese cross, shown in elevation at G, which rotates about the axis C. It is painted white, and the arms of the cross and the intervals are of equal width. It is illuminated by rays from S_1, which make an angle of 30° with its surface. DF is a white stationary screen and is illuminated by the rays from S_2, which likewise make an angle of 30° with its surface. The eye of the observer at E looks down a blackened tube and sees either the surface of the stationary screen or of the rotating cross, according as an open space or an arm of the latter comes into the line of view. Consequently if the illuminations of the two surfaces are unequal the observer experiences a sensation of flicker. If the distances of the sources are adjusted so as to make the illuminations equal, then the sensation of flicker disappears, and the ratio of the intensities of the sources can be obtained by the inverse square law.

The importance of the flicker photometer lies in the fact, that if the illuminations have different colours as well as different intensities, the cross can be rotated at such a speed that the colours fuse, resulting in a field of uniform hue, but yet the intensities do not, and so the sense of flicker remains. The instrument is thus admirably adapted for heterochromatic photometry, and there are one or two different types of it on the market. The rotation is usually done by clockwork.

The flicker photometer does not give the same results as the other methods at low intensities, when the comparison is influenced by the Purkinje effect.

The Relative Sensitiveness of the Eye to Light of Different Colours. The luminosity curve shown on p. 342 was for an electric arc, and gives the relative brightness of the different colours in its spectrum. Unfortunately the distribution of energy in the spectrum of that arc is not known, so that Abney's results do not inform us as to the relative sensitiveness of the eye to the different colours of the spectrum. This information has been obtained in a very thorough manner by a recent investigation due to H. E. Ives.

The apparatus is represented in fig. 284. It consisted of a modified Hilger constant deviation spectroscope, and employed the principle of the flicker photometer. As source of light a tungsten lamp was used at S_1 with a frosted glass between it and the slit. F was a rotating sector which took the place of the Maltese cross in fig. 283. Its surface was illuminated by a carbon lamp S_2 which could be moved to and fro along a scale. The eyepiece of the telescope was removed, and in its place a diaphragm with a circular aperture in its centre was placed in the focal plane of the object glass. The eye of the observer looking in at this aperture saw either the surface of the prism P, which appeared illuminated by whatever colour of the spectrum happened to fall on E, or instead of this he saw the surface of the disc F which was illuminated by white light. S_2 was kept fixed and the brightness of the part of the spectrum in question measured by altering the width of the slit until the sensation of flicker disappeared. The brightness was inversely proportional to the width of the slit. Thus a luminosity curve of the whole spectrum was obtained. The distribution of the energy in the spectrum of the lamp S_1 was known from previous

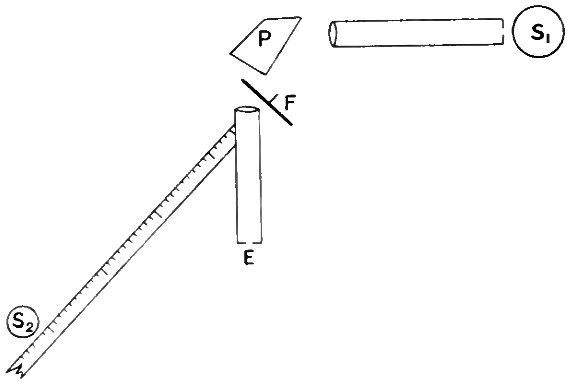


FIG. 284.

experiments, and by dividing the number that represented the energy at any wave-length into the number that represented the luminosity the curve shown in fig. 285 was obtained. It gives the mean of the results for eighteen observers, and is for an illumination of 25 metre-candles. The curve represents the sensitiveness of the eye for light

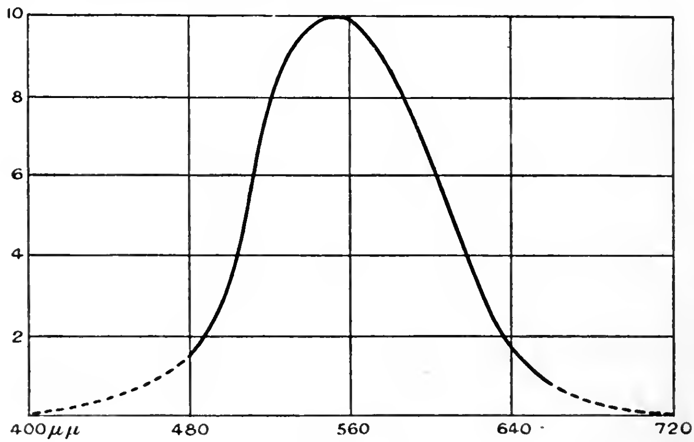


FIG. 285.

of different wave-lengths. We learn from it, for example, that if the same quantities of energy are in succession changed into light of the wave-lengths 6400 A.U., 6000 A.U., and 5600 A.U., the effect on the retina will be in the ratio of the numbers 1·7, 6·4, and 9·8. The curve has a smooth symmetrical shape. Indeed I find it hard to believe that such a curve is due to the superposition of the effects of *three* independent sets of nerves.

The advantages of Ives' method over the colour patch apparatus are that a tungsten lamp forms a steadier and more satisfactory source of light than an electric arc, and that at the intensities used the flicker photometer gives much more consistent results than are obtained by merely placing a white surface and a coloured surface side by side and matching their brightness. The colour patch apparatus requires a very bright source of light and could not be used with a tungsten glow lamp.

Haidinger's Brushes. If we look at a bright cloud through a nicol, and revolve the latter round its axis, so as to rotate the plane of polarisation, faint blue and yellow brushes appear at the point in the sky towards which the gaze is directed, and rotate with the nicol. They disappear gradually if the rotation stops, but become visible if the nicol is again rotated. They are known as Haidinger's brushes. Helmholtz explained them by assuming that the yellow spot in the eye is double refracting, and that in the case of blue light the extra-ordinary ray is absorbed more strongly than the ordinary ray.

EXAMPLES.

(1) Why can near objects not be seen distinctly when the eye is immersed in water?

(2) Determine the near and far points of your own eyes, and calculate their amplitude of accommodation. Compare the result with that given by the table on p. 337.

(3) Rule two parallel black lines close together on a piece of white card-board and measure the distance at which you can just see them separate. Hence calculate the angle which they subtend at the eye, and compare your result with the value given for the normal eye, namely 1′.

(4) The far point of a certain shortsighted person is situated at a distance of 30 cms. from the eye. Calculate in diopters the power of a spectacle lens which will enable him to see objects at a distance.

CHAPTER XX.*

LAMPS AND ILLUMINATION.

Flames. Flames are the simplest and historically the first means of producing light. As different stages in their development we have the primitive pine torch, the oil lamp of the ancients, the rush-light of the middle ages, the different kinds of candle from the tallow dip to the modern paraffin candle, the burning whale, seal, or bear fat of the Esquimaux, the Argand lamp, ordinary coal gas and acetylene gas. The action in all these cases is fundamentally the same—hydrocarbons are burned in air—and the spectrum is in every case the continuous spectrum of the incandescent carbon particles with the discontinuous spectra of water vapour and carbon dioxide and other gaseous products of combustion superimposed upon it.

As typical of the various flame illuminants we may consider the case of the ordinary paraffin candle. In the candle the flame melts the wax at the foot of the wick forming a cup there. The melted wax is sucked up the wick by capillary attraction and vaporised and burned as a vapour in the flame. The upward draught of air to the flame keeps the sides of the candle cool and prevents the edge from melting. The wax of course prevents the flame from burning too far down the wick. The top of the wick bends over into the edge of the flame and is consumed there.

The transparent region of the candle flame which surrounds the wick consists of unburnt vapour. Below this the flame has a blue edge. This blue edge gives the Swan spectrum, and it is not known yet whether the latter is due to carbon dioxide or monoxide. Above and surrounding the transparent region is an opaque luminous sheath which gives by far the greater part of the light of the flame. Its light is due to incandescent solid carbon particles which are formed by the decomposition of the vapour and have not yet combined with the oxygen of the atmosphere to form carbon dioxide. In the bunsen flame the oxygen combines with the carbon before it has time to radiate; if there is not enough oxygen to combine with the carbon the flame smokes. The bright opaque luminous sheath that gives the chief light of the flame is surrounded by a faint luminous mantle, which is difficult to see.

The chief thing that concerns us about a source of light is its effici-

* Much of the matter of this chapter is taken from the author's "Studies in Light Production" ("The Electrician" Co.).

ency. Suppose R is the total quantity of energy radiated by the source—this includes infra-red, visible and ultra-violet—and that L is the total quantity of light radiated per second. For this purpose we can define light simply as radiant energy, the wave-length of which lies between 4000 A.U. and 7600 A.U. Then L/R is defined as the radiant efficiency of the source. Besides the loss due to radiation the source also loses energy by conduction and convection. If we take this loss into consideration and let Q denote the total quantity of energy consumed by the source per second, then L/Q is defined as the luminous efficiency of the source. Since Q includes R, the luminous efficiency is always less than the radiant efficiency.

The old way of determining the radiant efficiency of a light source was to let the rays fall upon a thermopile and note the galvanometer deflection. The latter was proportional to R. Then a glass cell containing an aqueous solution of alum was placed in the path of the rays and the deflection again noted. The energy of the ultra-violet radiations was supposed to be negligible, which is nearly always the case, and the alum solution was supposed to absorb all the infra-red rays but let the light rays pass. Hence, when a correction was made for the loss of light by reflection, the second deflection should be proportional to L, and thus the ratio L/R could be determined.

The disadvantage of this method is, that a solution of alum lets a considerable proportion of the infra-red rays pass. The result thus always appears too high. An aqueous solution of ferrous ammonium sulphate is a better filter for keeping out the dark heat rays than alum. Despite a widespread impression to the contrary, a solution of alum is no better for this purpose than water itself.

A more satisfactory way is to employ a spectroscope equipped with a thermopile, and determine the distribution of the energy in the spectrum of the light after it comes through a water filter. Suppose fig. 286 represents the galvanometer deflections plotted against the wave-length, and the dotted line gives the end of the visible spectrum. Then the area of the whole curve is proportional to the radiation that comes through the filter, and the area of the part to the left of the line to the fraction of that radiation that is light. Hence the light can be separated from the dark heat completely.

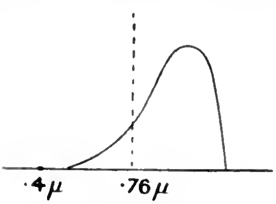


FIG. 286.

The radiant efficiency of the acetylene flame is 0·039. The value for a candle flame will be practically the same as the value for the Hefner lamp, which has been found to be 0·0096.

We can calculate the luminous efficiency of the Hefner lamp in the following way: The total radiation from a Hefner lamp as determined by Ångström with his pyrheliometer is $2{\cdot}15\ 10^{-5}$ cals. per sq. cm. per sec. at a distance of a metre. Hence in a second the flame radiates

$4\pi \times 10^4 \times 2{\cdot}15\ \ 10^{-5} = 2{\cdot}7$ cals. The quantity of amyl acetate consumed by the Hefner lamp in 1 sec. is 0·002678 gms., and the heat given out by the combustion of 1 gm. is 7971·2 cals., consequently the heat given out by the lamp per sec. is $7971{\cdot}2 \times 0{\cdot}002678 = 21{\cdot}35$ cals. The ratio of energy radiated to energy consumed is therefore only $2{\cdot}7/21{\cdot}35 = {\cdot}126$, the greater part of the energy consumed being lost by convection and conduction. Hence since the radiant efficiency of the Hefner lamp is 0·0096, its luminous efficiency must be ${\cdot}0096 \times {\cdot}126 = {\cdot}0012$. The efficiency of the Hefner lamp as an energy transformer is consequently not much better than $\frac{1}{10}$ per cent. No flame has a luminous efficiency much greater than this.

The Welsbach Mantle. The principle of this light is that a mantle of organic material is impregnated with a mixture of thorium oxide and cerium oxide in the proportion of about 1 part of the latter to 99 parts of the former. As soon as the mantle is fixed in position the organic material is burned off. The oxides remain and give out a bright light when they are heated in the colourless bunsen flame. The proportions of the oxides must be nearly right, otherwise the light is much dimmer. The inventor, Auer von Welsbach, did not come upon the discovery of the mantle suddenly but as a result of systematic experiments with different substances.

There was some doubt at first as to why the Welsbach mixture gave such a bright light, but, when the matter was fully investigated, it was found to be due solely to the energy spectrum of the mixture itself. The mantle is in a state of thermal equilibrium. It gains heat from the flame and loses it by conduction, convection, and radiation. Now the respective values of these losses are approximately the same, no matter what the mantle is impregnated with, but the proportion of the radiation loss falling in the visible spectrum differs from substance to substance. The Welsbach mixture is distinguished by the fact, that it possesses what is known as selective radiation, that a very large proportion of its radiation falls in the visible spectrum in the green and yellow. The radiant efficiency of the mantle is only 2 per cent, but our definition of radiant efficiency does not take account of what part of the visible spectrum the energy falls in.

Carbon Glow Lamps. The carbon glow lamp has been in use for more than thirty years. Its invention was due principally to the labours of Thomas Alva Edison and J. W. Swan. It consists of a carbon filament inside an exhausted glass bulb. The filament is heated to incandescence by the passage of an electric current. A platinum wire was tried at first instead of the filament but was not a success, for the platinum had to be raised nearly to its melting-point in order to give light economically, and an accidental variation of the current was then sufficient to cause the wire to burn through.

The bulb of the glow lamp is exhausted for two reasons, first to prevent chemical action between the filament and the air, and second to prevent loss of heat by conduction and convection from the filament to the bulb and consequently to the outer atmosphere. If the lamp is filled with hydrogen or nitrogen, gases which do not react with the filament, much more energy must be supplied to the lamp in order that it may give the same amount of light.

Carbon filaments are prepared by squirting a viscous solution of cellulose through a fine hole, so as to form it into a thread, and then carbonising this thread. The carbonised thread is then subjected to a process called flashing, which consists in raising it to incandescence in an atmosphere of hydrocarbon vapour ; the heated filament decomposes the gas, and a hard lustrous coating of graphite is deposited on the surface of the filament. Flashing improves the durability of the filament and renders its cross-section and conductivity uniform. Carbon possesses two advantages as a material for filaments, namely, its ability to stand high temperatures and its high specific resistance. It possesses the disadvantage, that after a time a black deposit forms on the inside of the bulb due to the evaporisation of carbon from the filament, and this black deposit diminishes the candle-power of the lamp. When the latter has diminished to about 80 per cent of its initial value, it is cheaper to replace the bulb by a new one. The time required for this diminution is said to constitute the useful life of the lamp and is roughly about 800 hours.

A 16 candle-power carbon lamp on a 250 volt circuit takes about ·2 amps. It requires therefore $250 \times \cdot 2 = 50$ watts, that is about $50/16 = 3{\cdot}1$ watts per candle. This is how practical men are accustomed to specify the performance of a lamp, in watts per candle, and the value given for carbon glow lamps is from 3 to 3·5.

Metal Filament Lamps. If the voltage on a carbon glow lamp is increased much above its proper value, the temperature of the filament is increased, and the lamp gives much more light for the same energy, but is soon burnt through. This led to a search for a filament that would stand a higher temperature than carbon without burning through, and in 1902 a lamp employing osmium as a material for the filament and using 1½ watts per candle was put successfully on the market. It was succeeded by the tantalum filament lamp in 1905, and the latter was immediately followed by the tungsten filament lamp, which latter under the various names of Osram, Mazda, etc., at present holds the field. The osmium lamp is no longer made, as it is less efficient than the tungsten lamp.

The tungsten filaments were at first made from a paste by squirting the latter through a small hole in a diamond under high pressure, and were then very easily broken. They are now, like the tantalum filaments, wire-drawn and consequently much stronger. The specific resistance of tantalum and tungsten is very much less than the specific resistance of carbon, consequently the filaments have to be much thinner, the diameter having, for example, the value ·03 mm. There is also another difference ; the resistance of tantalum and tungsten increases with the temperature, while the resistance of carbon possesses a negative temperature coefficient and decreases with rise of temperature and increase of current. Tungsten lamps are recognised as requiring 1 or $1\frac{1}{4}$ watts per candle. Their superiority over the carbon lamp is due partly to their higher temperature and partly to selective radiation, to the fact that they radiate a smaller proportion of infra-red dark heat.

The following table gives some figures for different lamps obtained by G. Leimbach. The first column gives the name of the lamp, the second the number of watts required per British candle, the third the ratio of total energy radiated to total energy supplied, the fourth the radiant efficiency, and the fifth the luminous efficiency. The last two lamps on the list are tungsten lamps. The numbers in the second column are higher than the usually accepted values :—

1	2	3	4	5
	Per cent.	Per cent.	Per cent.	Per cent.
Carbon glow lamp . . .	4·2	61·9	2·85	1·75
Nernst ,, ...	2·2	49·2	4·43	2·17
Tantalum ,, ...	2·2	64·8	4·26	2·75
Osram ,, ...	1·7	75·6	4·63	3·50
A.E.G. metal filament lamp	1·9	80·5	4·41	3·55

The Nernst Lamp. The filament of this lamp is a combination of oxides of cerium, thorium, and zirconium. It does not react chemically with the atmosphere, and hence does not require to be enclosed in a vacuum. The Nernst filament or glower, as it is called, does not conduct at low temperatures and must be warmed before the lamp is started. In the first lamps this warming was done with a match, but afterwards the lamps were provided with an electric heater, which cut out automatically whenever the lamp started. As the resistance of the glower decreases rapidly with rise of temperature it must be provided with a ballast resistance. The conduction in the Nernst filament is electrolytic, and its light has a reddish tint. It has never been widely used in this country principally owing to the advent of the metal filament lamp.

Distribution Curve of an Illuminant. No sources radiate light equally in all directions. When the intensity of a source is compared with the intensity of a standard such as the pentane standard or the Hefner lamp, it is always the light radiated by the latter in a horizontal direction that is used. In specifying the performance of an illuminant it is necessary not only to give its candle-power, but also the directions in which it sends out most light. This is done by a distribution diagram. Fig. 287, for example, gives such a diagram for a bare inverted incandescent mantle. In the figure the radii give candle-power and the angles directions. Thus we see that in a downward direction at an angle of 45° with the vertical the intensity of the source is about

70 candles. Owing to the burner getting in the way little light goes in an upward direction. Of course it may happen that the distribution

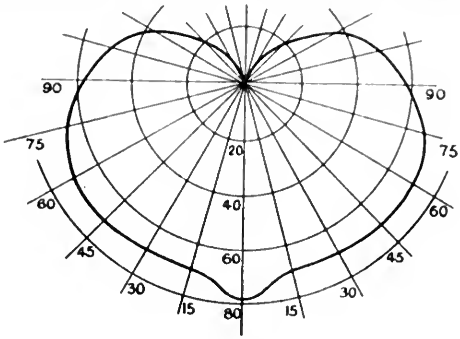


FIG. 287.

of light is different in different vertical planes, in which case, unless specially designated, the distribution curve is taken to represent the mean of all the distribution curves in the different vertical planes.

There is a widespread impression that the area of the distribution curve is proportional to the total quantity of light emitted by the source. This is not the case, but if the distribution curve is rotated about its axis, then it is easy to see that the volume of the solid thus formed is proportional to the total quantity of light emitted by the source. The true measure of the performance of a source is not its intensity in any particular direction but its mean spherical candle-power, that is, the intensity which it would have if the light it actually does radiate were radiated equally in all directions.

Determination of Mean Spherical Candle-power from Distribution Curve. As the distribution curve has never an exact mathematical form, this has to be done approximately.

Let I be the intensity in a direction θ, θ being measured from the horizontal. Divide the whole surface formed by rotating the distribution curve into n horizontal zones of angular width π/n, and suppose that the value of I is constant throughout each zone. The solid angle subtended by a zone is equal to the area which the radii from the origin to the edges of the zone intercept on a sphere of radius unity. The area intercepted on the sphere is equal to the length of the arc multiplied by the mean circumference, or

$$\frac{\pi}{n} 2\pi \cos \theta,$$

where θ gives the position of the mean circumference. The radiation through the zone multiplied by the solid angle subtended by the zone is consequently

$$\frac{2\pi^2 I \cos \theta}{n}.$$

Calculate this expression for each of the n zones, take the sum and divide by 4π, the solid angle subtended by all the n zones. The result

$$\frac{1}{4\pi}\Sigma\frac{2\pi^2 \mathrm{I} \cos\theta}{n} \quad \text{or} \quad \frac{\pi}{2n}\Sigma \mathrm{I} \cos\theta$$

is the mean spherical candle-power.

The calculation is much facilitated by always using the same zones and working in tabular form.

A distribution curve can be obtained step by step on an ordinary photometric bench by gradually inclining the lamp and so getting the intensity at different angles. There are, however, instruments called integrating photometers, which by means of reflectors placed round the source give the mean spherical candle-power at one operation.

Subsidiary Standards. In the photometry of artificial illuminants it is not usual to compare them directly with primary standards such as the pentane or Hefner lamps, as these latter are not convenient for rapid operation, but we employ instead intermediate standards which have been calibrated in terms of a primary standard. Carbon or tungsten glow lamps are used as such intermediate standards, but they have first to be seasoned by running them for 100 hours, as in this first period of their use the intensity varies irregularly. Also when they have been used as standards for 100 hours or thereabouts they ought to be checked with a primary standard. A 5 per cent variation from the normal voltage means 28 per cent variation in the candle-power in the case of a carbon glow lamp and 18 per cent variation in the case of a tungsten lamp. As the voltage of a city lighting circuit varies considerably about its normal value, glow lamps when used as standards should be run off a storage battery.

The Arc Lamp. The arc is the oldest electrical illuminant. Formerly only the arc passing between carbon electrodes was used for lighting purposes. These electrodes are made from a mixture of powdered coke and pitch, which is moulded under heavy pressure and then baked at a high temperature to drive off volatile matter and harden them. To start the arc it is " struck," i.e. the two electrodes or carbons which are at a difference of potential of 70 or 80 volts are suddenly brought into contact and then pulled apart. This causes an electric discharge characterised by intense light to pass between them.

After the discharge has passed some time, the end of the positive carbon or anode becomes hollowed out and the end of the negative carbon or kathode becomes pointed. Both the carbons are consumed by the arc, the positive one twice as fast as the negative one. There is a luminous vapour between the electrodes, and the ends of both carbons give off a bright light, 85 per cent of the total light coming from the hollow or crater on the positive carbon. If the current is too large, the crater becomes too large to occupy the end of the carbon and the lamp begins to hiss. The carbons of the arc lamps used for street illumination strike and feed together automatically, but in the arcs used for projection work this is usually done by hand.

In recent years the arc has been improved in different directions. It has been enclosed in a tightly fitting glass globe so as to restrict the supply of air and lengthen the life of the carbons. Then, instead of mounting the carbons coaxially, they have been inclined to one another with the anode pointing downward; the crater can then radiate freely without being obscured by the negative carbon. Also the electrodes have been impregnated with salts of strontium, calcium, barium, and titanium. This has the effect of making the luminous vapour between the electrodes develop into a brilliant flame, yellow in the case of calcium salts, red in the case of strontium salts, and white in the case of barium and titanium salts; this flame gives the greater portion of the light. The flame arc is the most efficient of all illuminants, requiring only $\frac{1}{4}$ watt per candle.

Future Progress. The development of illumination within the past thirty years has been to some extent a duel between gas and electricity. The gas flame was giving place to the carbon glow lamp, when the situation was changed by the invention of the Welsbach mantle. The glow lamp had the advantage over the first incandescent mantles that it shed its rays down instead of up, but the invention of the inverted mantle made both illuminants equal in this respect. The development of flame arc lamp lighting was met by high pressure gas lighting, i.e. the use of very large mantles, which are heated to higher temperatures than the ordinary mantle by having the gas at a higher pressure.

It is customary to classify sources of light under two headings, namely, temperature radiation and luminescence. In cases of temperature radiation the light is emitted solely as a result of the body being heated, the electrons which emit the light waves being caused to vibrate as a result of the motion of the molecule as a whole. In cases of luminescence electric energy or chemical energy is changed directly into the energy of light waves without passing through the intermediate stage of heat; the electrons are set vibrating without kinetic energy being given to the molecules as a whole.

All flames, the carbon and metal filament lamps, the incandescent mantle and the carbon arc are cases of temperature radiation. The metal filament lamps and particularly the incandescent mantle are distinguished by selective radiation, i.e. an unusually large proportion of their radiation falls within the visible spectrum, but the advance of each of the above sources over its predecessors lies chiefly in the employment of a higher temperature. As the temperature is increased selective radiation tends to become less, also the highest attainable temperatures are practically reached, so that not much further progress is to be expected from temperature radiation.

The only case of luminescence mentioned in this chapter has been the flame arc, although other cases, the vacuum tube and the mercury arc, have been described in Chapter XIV. Both the latter have been used commercially as methods of illumination. McFarlan Moore employed vacuum tubes 40 to 220 feet long and about 2 in. in diameter with a voltage of about 10,000 to 12,000 produced by a transformer. The gas used was nitrogen and the current in the tube was 0·3 ampere. Advantages of the Moore vacuum tube are its low intrinsic brightness, it gives a soft light and there are no brilliant filaments or electrodes to hurt the eye. Also, owing to its length, it gives a very uniform illumination; there are no sharply defined shadows such as are produced by point sources. With vacuum tubes the colour of daylight can also be imitated very accurately. The efficiency of the Moore tube, according to tests by different authorities, is 1·78 or 1·53 watts per candle. It seems, however, that if the gas employed is neon, the consumption of watts per candle is less than this. The objection to the Moore tube is that it is revolutionary.

The most successful mercury arc, the Cooper-Hewitt tube, has been described on p. 250. Its efficiency is roughly $\frac{1}{2}$ watt per candle, and its disadvantage is the strong green colour of its rays, which give the colours of the objects they illuminate a false value; thus they make yellow hair appear green and reds a bluish-brown. If the Cooper-Hewitt tube is used together with carbon glow lamps or a red fluorescent reflector, the colour is improved, but at the same time the efficiency is diminished. The best results have been attained with quartz tubes, which stand a higher current than glass. The higher current increases the pressure, gives the spectrum a continuous background, and so improves the colour of the lamp; at the same time it increases its efficiency to $\frac{1}{4}$ watt per candle.

Luminescence is not nearly so well understood as temperature radiation, and the vacuum tube has not been very thoroughly studied from the point of view of an illuminant. By putting a bolometer inside a vacuum tube R. W. Wood showed that the temperature was about 30° or 40° C. In the case of the vacuum tube we are therefore dealing with something quite different from temperature radiation. If any very great progress is to be made in the future, it will be made from the side of luminescence.

Diffusing Globes. Through the experience of ages the eye has become accustomed to light coming obliquely from above. Intense light coming from any other direction, for example, sunshine on snow, produces irritation and discomfort. When the eye is exposed to a bright light the diameter of the pupil contracts automatically, but this power of accommodation is limited. The intrinsic brilliancy of the crater of the arc is 200,000 candles/sq. inch and of a tungsten lamp filament 1000 candles / sq. inch; an incandescent gas mantle gives 20 or 25 candles/sq. inch. These values are all too high for comfort, so these sources must all be used with diffusing globes so as to bring the intrinsic brilliancy down. The frosted incandescent lamp has an intrinsic brilliancy of approximately 3 or 4 candles per square inch. Diffusing globes act either by having their surfaces sand-blasted or etched, or owing to fine particles held in suspension in the glass, or owing to their surfaces being ribbed, so that they refract the rays like prisms of variable angle. Diffusing glasses do not scatter the transmitted light equally in all directions but principally in the direction of the incident light. Since the light sources are always placed above the heads of the people in a room, it is necessary that the globes should be constructed so as to send the greater proportion of light down. Instead of using globes and reflectors the same end can be attained by having naked lights and the ceiling and upper portions of the walls white, and screening the lights so that their direct rays do not reach the eyes; then the white walls and ceiling act as a reflector.

Diffuse Reflection. The degree of illumination of a room depends not only on the light sources but on the colour of the walls. All walls reflect the incident light diffusely.

Let MS be a section of a wall, and let i be the angle which the incident light makes with PN, the normal at P. Let I be the intensity of the incident light, i.e. the quantity of energy received per sq. cm. per sec. Then if a straight line PQ be drawn of length r, making an angle e with PN and not necessarily in the plane NPM, and an area dS be isolated on the surface of the wall at P, the intensity of the radiation from this area received at Q is

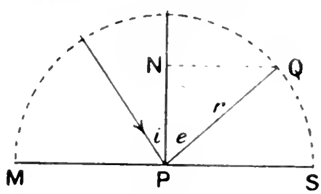


FIG. 288.

$$\frac{c\text{I}d\text{S} \cos i \cos e}{r^2},$$

where c is a constant. This expression is known as Lambert's cosine law of diffuse reflection.

Let us find the total quantity of light radiated out by the area dS in unit time. Suppose the angle e increases by de and the point Q consequently moves to Q′; then QQ′ $= rde$. Suppose now that QQ′ is rotated round PN as axis. The area swept out is

$$2\pi\text{NQ} \times \text{QQ}' = 2\pi r^2 \sin e \, de.$$

The total light received by this area is

$$\frac{c\text{I}d\text{S} \cos i \cos e}{r^2} \times 2\pi r^2 \sin e \, de = \pi c\text{I}d\text{S} \cos i \times 2 \sin e \cos e \, de$$

$$= \pi c\text{I}d\text{S} \cos i \sin 2 e \, de.$$

The total light radiated out by dS in unit time is consequently

$$\pi c\text{I}d\text{S} \cos i \int_0^{\frac{\pi}{2}} \sin 2 e \, de = \pi c\text{I}d\text{S} \cos i,$$

and the light received by dS is IdS $\cos i$. The ratio of the light radiated to the light received, or the albedo, as it is called in astronomy, is consequently

$$\pi c.$$

Some values of this quantity for ordinary daylight are given in the following table :—

White cartridge paper	·80
Ordinary foolscap	·70
Yellow wallpaper	·40
Emerald green paper	·18
Black cloth	·012
Black velvet	·004

It is obvious from the table that gas and electricity can be saved by choosing a light wallpaper.

Lambert's cosine law is obeyed very well by a scraped surface of plaster of Paris, but with varying degrees of approximation in the case of other surfaces. The usual case is to have some regular reflection and diffuse reflection occurring simultaneously such as, for example, in the case of glazed paper.

According to Lambert's law the quantity of light received at Q, per unit of area at right angles to PQ, is

$$\frac{cI\,dS \cos i \cos e}{r^2}.$$

Owing to its being seen obliquely the surface dS appears at Q to have the area $dS \cos e$. The surface dS has thus the apparent brightness

$$\frac{cI \cos i}{r^2},$$

which is consequently independent of the angle e. This is the physical basis on which Lambert's law rests. If the eye is moved round the dotted circle in fig. 288, dS appears always equally bright.

A luminous sphere, for example, a red-hot copper sphere heated by holding it in tongs over a bunsen flame, appears as a disc of uniform brightness. The apparent brightness of the different elements on its surface is the same, no matter what angle the normal to the element makes with the straight line joining the latter to the eye. The radiation from an element of the surface in a direction making an angle e with the normal to that element must consequently be proportional to $\cos e$.

Illumination Photometry. The photometers described in Chapter XVIII are used for measuring the intensity of a light source in the laboratory. It is often necessary to measure in foot-candles or metre-candles the degree of illumination of a surface in a room or in the street, and for this purpose instruments called illuminometers have been devised. The foot-candle is the illumination received on a surface facing a light of 1 candle-power at a distance of 1 foot, and the metre-candle the illumination received on a surface facing a light of 1 candle-power at a distance of 1 metre. Illuminometers usually contain a small electric lamp which is run off accumulators. This lamp illuminates a matt surface inside a dark chamber, and the brilliancy of this matt surface can be regulated by means of an adjustable diaphragm or otherwise. The eye looking into the instrument sees through a hole close beside this comparison surface the surface whose illumination is to be measured, and the comparison surface is adjusted to the same degree of illumination by means of the adjustable diaphragm or by the insertion of smoke glasses. The instrument is calibrated by the use of surfaces of known illumination. The degree of accuracy aimed at in the use of such instruments is not high.

The illumination of an object in ordinary daylight is roughly 1000 foot-candles. From 1 to 4 foot-candles is sufficient for office desks, schoolrooms, etc., and probably 0·1 foot-candle would be the average illumination in a well-lighted street at night. Of course in illuminating a street it is a mathematical problem in itself to choose the distance apart of the lamps and their height above the ground, so as to combine uniformity of illumination with brilliancy. If the lamps are too high, the illumination on the reference plane, which is usually taken as 4 feet above the level of the street, will be uniform but weak; if they are too low it will be very unequal.

EXAMPLES.

(1) Determine the distribution curve of a carbon glow lamp on an ordinary photometer bench. The lamp can be held in a clamp and its inclination measured with protractors.

(2) Calculate the mean spherical candle-power of the same lamp by the method given on p. 357.

(3) Calculate the mean hemispherical candle-power of the same lamp. (Proceed as in example immediately above, but use only the zones on the hemisphere which does not contain the holder.)

(4) Measure experimentally the amount of light reflected diffusely in different directions by a square of white cardboard of 2 inches side. (Use the square of white cardboard in place of one of the lamps on the ordinary photometer bench and balance it against a weak source. Illuminate it by a beam from an arc lamp used for projection; remove the projecting lens and have the cardboard beyond the image of the crater formed by the condenser, and far enough into the diverging cone to prevent the illumination changing owing to the arc wandering. Stray light must be very carefully guarded against and readings taken rapidly and often.)

(5) Measure in foot-candles the illumination in a room on a plane 2½ feet above the floor. Improvise an illuminometer for the purpose, using as comparison source a four volt lamp.

(6) Determine the candle-power and calculate the intrinsic brilliancy in candles per sq. inch for different light sources with and without diffusing globes.

(7) Invert a carbon glow lamp, dip it into a glass beaker containing clean water and note the rise of temperature in five minutes. Repeat the experiment but mix ink with the water so that the lamp cannot be seen through it. In the first case the water is heated by all the heat emitted except the visible and near infra-red radiation; in the second case these are absorbed also. Hence the difference of the rises divided by the larger rise gives a rough measure of the luminous efficiency. (The experiment does not give definite results unless very carefully carried out. Criticise it from the point of view of the light internally reflected in the beaker.)

PART IV.

MATHEMATICAL THEORY.

CHAPTER XXI.

THE NATURE OF LIGHT.*

PREVIOUS to Newton's experiments it was thought that the prism made the spectrum. According to Newton, however, it merely separated the different colours out, and white light consisted of the different colours superimposed. Both of these statements are, in a sense, true, but their full meaning is not easily grasped. The simplest way of making the matter clear is to study first a similar problem in hydrodynamics which lends itself more easily to treatment. Then the optical problem can be solved by analogy.

Hydrodynamical Analogue. Let us consider a sheet of water of unlimited depth. Take the Ox axis in its undisturbed surface, measure the Oy axis vertically downwards, and let the coordinates x, y refer to points on the surface of the water. Let the surface of the water be given initially by

$$y = \frac{\cos \frac{5}{2} \tan^{-1} (x/h)}{(x^2 + h^2)^{5/4}}.$$

This curve is represented in fig. 289. It has a depression at $x = 0$ given by

$$y = \frac{1}{h^{5/2}}.$$

On both sides of this depression it ascends rapidly to a point above the x axis and then descends again towards the axis, at first rapidly but then more slowly, and finally reaches it at $x = \pm \infty$. When x is greater than $10h$ the curve practically coincides with Ox. Let us suppose that the depression is the same for all the points for which x is the same. The initial form of the surface could thus be produced by waiting until it was perfectly plane and then laying a very long properly curved rod down upon it.

* The matter of this chapter is not usually given a prominent position in books on light, but it seems to deserve it owing to the interest taken at present in the quantum theory. Many writers on Radiation seem unaware that the orthodox theory permits of light being propagated in pulses, and of the dispersion and interference of such pulses.

The mode of treatment here follows closely two papers by the Author, "Proc. Roy. Soc.," A **89**, p. 399 (1913), and A **90**, p. 288 (1914). From the nature of the subject the chapter is a difficult one; it may be omitted without prejudice to those that follow.

What happens if the rod is raised very rapidly? This problem has received a rigorous solution. It is found that two waves travel out

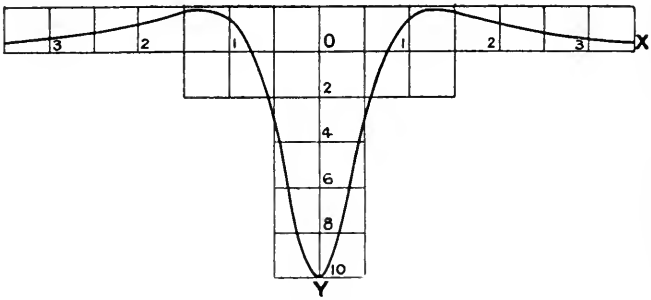


FIG. 289.

with equal velocities from the initial depression, one in the direction of $+x$ and the other in the direction of $-x$, and that, as these waves progress, their form alters. From being single waves they gradually develop into groups. When the two waves are still close to the origin their shape has a somewhat complicated mathematical expression; when they get far out, this expression becomes extremely simple. It is then given by

$$\frac{t^4}{x^{9/2}} e^{-t^2h/4cx^2} \cos \frac{t^2}{4cx}.$$

t is measured from the time of the initial disturbance; c is a constant depending on gravity. The expression has also a constant coefficient which has been omitted here in the interests of simplicity.

Let us graph the above solution as a function of x for definite values of t, c, and h. Take, for example, $t^2/c = 540$ and $h = \frac{1}{15}$. Then the expression becomes, on omitting the constant coefficient,

$$\frac{1}{x^{9/2}} e^{-9/x^2} \cos \frac{135}{x}.$$

The first factor, $\frac{1}{x^{9/2}}$, is infinite when x is 0, and 0 when x is infinite; the second factor, e^{-9/x^2}, is 0 when $x = 0$, and 1 when x is infinite. The product is 0 for x infinite and at first indeterminate for $x = 0$, but on investigation it is found to be zero for this limit also; it is represented by the dotted line in fig. 290. For reasons which will be obvious

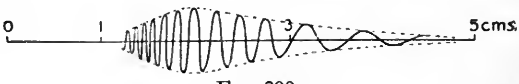


FIG. 290.

immediately the dotted line has been drawn on both sides of the axis. It has not been drawn for the region 0 to 1, because, as has been mentioned above, the solution is not accurate near the origin.

Let us now consider the last factor $\cos 135/x$. It is equal to 1 when $135/x = 2n\pi$, n being a positive integer. The values of x between 1 and 5 which satisfy this equation are 1·02, 1·07, 1·13, 1·19, 1·26, 1·34, 1·43, 1·54, 1·66, 1·79, 1·95, 2·15, 2·39, 2·68, 3·06, 3·58, 4·29. Consequently at these points the cosine factor does not affect the value of the product of the first two factors and the curve representing the complete expression touches the upper dotted line. At points between these values it touches the lower dotted line. Hence the complete expression is represented by the full line.

The full line represents a group of waves in which the wave-length gradually decreases from the front to the rear of the group.

§ Let us go back now to the general expression and no longer restrict ourselves to a single value of t. The dotted curve is given by

$$y = \frac{t^4}{x^{9/2}} e^{-t^2h/4cx^2}.$$

If we regard it as a function of x its maximum is given by

$$\frac{dy}{dx} = -\frac{9t^4}{2x^{11/2}} e^{-t^2h/4cx^2} + \frac{t^4}{x^{9/2}} e^{-t^2h/4cx^2} \frac{2t^2h}{4cx^3} = 0,$$

i e.

$$\frac{t^4}{x^{11/2}} e^{-t^2h/4cx^2} \left(-\frac{9}{2} + \frac{t^2h}{2cx^2}\right) = 0$$

which gives $x = t\sqrt{\frac{h}{9c}}$.

The maximum therefore moves forward with a uniform velocity $= \sqrt{h/9c}$. We can refer to this expression as the "group velocity". On substituting it for x/t the value of the maximum becomes

$$\frac{81c^2}{x^{\frac{1}{2}}h^2} e^{-9/4}$$

and consequently diminishes inversely as the root of x, as the group moves out.

Consider two points in the group exactly a wave-length apart; let x be the coordinate of one and $x + \lambda$ be the coordinate of the other. The phase $t^2/(4cx)$ must increase by 2π in moving from the one point to the other. Consequently

$$\frac{t^2}{4cx} - \frac{t^2}{4c(x+\lambda)} = 2\pi,$$

i.e.

$$\frac{t^2}{4c}\left(\frac{1}{x} - \frac{1}{x+\lambda}\right) = 2\pi,$$

or

$$\frac{t^2}{4c} \frac{\lambda}{x(x+\lambda)} = 2\pi,$$

and this gives, on neglecting λ in comparison with x,

$$\lambda = \frac{8\pi cx^2}{t^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

For a given time the wave-length increases from the rear to the front of the group; for a given point it decreases with time.

Let us fix our attention on a given point and suppose that the phase increases by 2π during the interval of time between t and $t + \tau$. Then τ denotes the period and is given by

$$\frac{(t+\tau)^2}{4cx} - \frac{t^2}{4cx} = 2\pi,$$

i.e.

$$\frac{2t\tau}{4cx} = 2\pi,$$

or

$$\tau = \frac{4\pi cx}{t}, \quad . \quad . \quad . \quad . \quad (2)$$

if τ be neglected in comparison with t. On eliminating x/t between (1) and (2) we obtain

$$\frac{\lambda}{8\pi c} = \left(\frac{\tau}{4\pi c}\right)^2 \text{ or } \lambda = \frac{\tau^2}{2\pi c}.$$

But λ/τ is the velocity of the individual wave at the point and time in question. Hence the wave velocity is given by $\tau/(2\pi c)$.

We shall now find the manner in which the energy is distributed over the different wave-lengths for a given value of t. We shall make the usual assumption that the energy is proportional to the square of the amplitude. Then if A is the amplitude at x, the energy contained in the part of the group between x and $x + dx$ is proportional to $\mathsf{A}^2 dx$; we assume here that dx contains a large number of wave-lengths. The change of wave-length in the distance dx is given by

$$d\lambda = d\left(\frac{8\pi cx^2}{t^2}\right) = \frac{16\pi cx dx}{t^2}.$$

The energy per wave-length therefore varies as

$$\mathsf{A}^2 \frac{dx}{d\lambda} = \frac{t^8}{x^9} e^{-t^2h/2cx^2} \frac{t^2}{16\pi cx}$$

$$= \left(\frac{t^2}{x^2}\right)^5 \frac{1}{16\pi c} e^{-t^2h/2cx^2}.$$

Substitute $t^2/x^2 = 8\pi c/\lambda$; then the energy per wave-length varies as

$$\frac{(8\pi c)^5}{\lambda^5} \frac{1}{16\pi c} e^{-h8\pi c/2c\lambda} = \frac{(8\pi c)^4}{2} \frac{1}{\lambda^5} e^{-4\pi h/\lambda}.$$

The energy belonging to the range of wave-lengths included between λ and $\lambda + d\lambda$ is consequently always the same.

§ We had for the maximum of the group $x = t\sqrt{h/9c}$. The wave-length is given by $\lambda = 8\pi cx^2/t^2$. Eliminate x/t between these two equations and $\lambda = \frac{8}{9}\pi h$. The result is independent of x and t and gives the wave-length at the maximum. This wave-length, which always rides at the top of the group, we shall refer to as the "dominant wave-length".

If we consider now the initial expression, it can be written in the following form :—

$$y = \frac{\cos \frac{5}{2} \tan^{-1}(x/h)}{h^{5/2}\{(x/h)^2 + 1\}^{5/4}} = \frac{1}{h^{5/2}} f\left(\frac{x}{h}\right).$$

It is obvious from the form of this expression that if h is halved in value, all the ordinates are increased in the ratio 1 to $2^{5/2}$ and moved in half way to the origin. The curve becomes narrower and sharper. We thus obtain the result, that the narrower and sharper the initial disturbance is, the shorter is the dominant wave-length.

We have considered the initial disturbance to be a depression, but can, of course, obtain the case of an elevation simply by changing the signs.

The reasoning on the preceding pages is somewhat close and so should be followed carefully step by step. If graphs are drawn wherever possible, the difficulties will disappear. The general results may be stated as follows:—

If on the surface of deep water an initial disturbance is applied along a straight line, then disturbances travel out in opposite directions on both sides of this straight line. These develop into groups of waves; as each group proceeds, the number of waves in it increases indefinitely and its shape alters. The velocity of the maximum of the group is constant. The range of wave-lengths included between λ and $\lambda + d\lambda$ has always the same velocity and carries always the same quantity of energy.

Another Way of Regarding the same Problem. We have by a well-known result (cf. Williamson's "Integral Calculus," § 124)

$$\frac{\cos \frac{5}{2} \tan^{-1}(x/h)}{(x^2 + h^2)^{5/4}} = \frac{4}{3\sqrt{\pi}} \int_0^\infty e^{-ha} \cos ax \, a^{3/2} \, da \quad . \quad . \quad (3)$$

The expression on the left gives the initial disturbance. The expression on the right is obtained by superimposing on one another an infinite number of cosine curves, the phases of which agree at the origin and the wave-lengths of which vary from ∞ to 0; the amplitudes of these curves are given by

$$\frac{4}{3\sqrt{\pi}} e^{-ha} a^{3/2},$$

where $a = 2\pi/\lambda$. It is more convenient here to express results in terms of a instead of λ. We shall refer to a as the parameter of the wave. If we consider the expression

$$e^{-ha} a^{3/2},$$

we find it has a maximum given by

$$\frac{d}{da} e^{-ha} a^{3/2} = -he^{-ha} a^{3/2} + e^{-ha} \tfrac{3}{2} a^{1/2} = 0,$$

or

$$a = \frac{3}{2h}.$$

On both sides of this maximum its value descends to 0, for at $a = 0$ it becomes 0, and at $a = \infty$ it takes at first the indeterminate form $0 \times \infty$, which, on applying the usual rule, becomes 0 also. The narrower the initial disturbance, the smaller is h and consequently the shorter the wave-length with the maximum amplitude.

The expression on the left-hand side of equation (3) has an appreciable value only in the neighbourhood of the origin, whereas each of the cosines on the right-hand side has the same value to infinity. Only at the origin, though, do their phases agree. In the neighbourhood of the origin they interfere partially. Elsewhere the interference is complete and hence the integral has the value zero.

It was shown on p. 370 that a wave with the period τ has the velocity $\tau/(2\pi c)$ and consequently the wave-length $\tau^2/(2\pi c)$. Hence $a = 2\pi/\lambda = 2\pi \times 2\pi c/\tau^2$. This gives $\tau = 2\pi\sqrt{c/a}$. Consequently a wave with the parameter a has the velocity

$$\frac{\tau}{2\pi c} = \frac{2\pi}{2\pi c}\sqrt{\frac{c}{a}} = \frac{1}{\sqrt{ac}}.$$

If we consider each cosine curve on the right of (3) to move with its own velocity and disregard the constant coefficient, then after time t the right-hand side has become

$$\int_0^\infty e^{-ha} \cos a\left(x - \frac{t}{\sqrt{ac}}\right) a^{3/2}\, da.$$

The phases of all the cosines no longer agree at the one point because their velocities are different.

The wave for which the parameter is $a + da$ has at the point x at the time t the phase

$$(a + da)\left(x - \frac{t}{\sqrt{ac}}\right) + a\frac{t}{2a^{3/2}c^{\frac{1}{2}}}da.$$

This differs from the phase of the wave with the parameter a by

$$\left(x - \frac{t}{\sqrt{ac}} + \frac{t}{2\sqrt{ac}}\right)da \quad \text{or} \quad \left(x - \frac{t}{2\sqrt{ac}}\right)da.$$

Hence if $x = \dfrac{t}{2\sqrt{ac}}$, the waves for which the parameter is included between a and $a + da$ have the same phase and reinforce one another. Thus the wave with parameter a is in evidence at that point but nowhere else. Owing to their different velocities the different waves separate and each shows at the point where it is reinforced by its neighbours, that is, where

$$a = \frac{t^2}{4cx^2} \quad \text{or} \quad \lambda = \frac{8\pi c x^2}{t^2}.$$

This is in accordance with the former result.

The cosine waves are not contained in the original disturbance in any physical sense. They are mathematical fictions used as a means of obtaining results. Each becomes real for the small fraction of its range in which it is reinforced by its neighbours.

Application to Light. It is the dispersive power of the medium that makes the different waves separate out in the preceding case, and, of course, in that case the dispersive power is a very great one. The velocity of a wave is $\tau/(2\pi c)$ and consequently can vary from 0 to ∞.

The velocity of light in interstellar space is constant and independent of the wave-length. If a light pulse comes from the sun, during its passage through interstellar space it preserves its form unaltered. As soon as it enters a dispersive medium such as the atmosphere or a plate of flint glass the form alters, the different wave-lengths separate out and the pulse tends to develop into a group. Since in air and glass the longer waves travel faster, just as in the preceding case of deep sea waves, so in air and glass the longer waves are in the front of the group. But in the preceding problem the dispersive power is much greater than that of air or glass; in the case of a group which has its dominant wave-length in the yellow it is roughly 14 times as great as the dispersive power of glass and 10^5 times as great as that of air.

It is easy to estimate the amount of dispersion that takes place when a pulse passes through a glass plate. If, for example, the plate is 1 cm. thick and the index of refraction is 1·6165 for the red hydrogen line and 1·6642 for the violet hydrogen line, typical values for dense flint glass, when the red hydrogen wave has reached the other side of the plate, the violet hydrogen component has travelled only $\frac{1{\cdot}6165}{1{\cdot}6642}$ cm. and has consequently a distance of $\frac{1{\cdot}6642 - 1{\cdot}6165}{1{\cdot}6642}$, or ·029 cm. still to go. The length of the train between these two points is consequently ·029 cm. This is 500 times the wave-length of sodium light, so that the passage through the slab changes the single pulse into a group of about 500 visible waves. After emerging into the air, owing to the low dispersion of the latter the form of the group changes extremely slowly.

Just as in the case of the water waves, the waves into which the incident light pulse can be resolved are not contained in that incident pulse in any physical sense. They are mathematical fictions and become real, each only for its own short range, after passing through the plate.

We shall next show how a single pulse is changed into a spectrum, first for the case of the diffraction grating and second for the case of the prism. The former is the easier of the two cases.

Action of a Diffraction Grating on a Light Pulse. Let a plane light pulse fall normally on a transmission grating. It falls on each transparent space simultaneously. By Huygens' principle each point on a transparent space can be regarded as a secondary source, but as the transparent spaces are small we can take all the points on each space together and represent them by an equivalent source at the centre of the space. Let A, B, C, and D be such equivalent sources. When the plane pulse falls on the grating, then secondary waves start out from A, B, C, and D. These secondary waves are represented in the diagram. They are of course still pulses; each consists of but a single maximum.

The four secondary waves superimpose approximately at E since the distance AD is small compared with the radius of the circles. Hence the telescope when placed in the position shown by the dotted lines at E receives a single pulse very similar to the original pulse, and this gives a white light image of the slit. If the telescope is placed at F it receives four separate waves. In the diagram the fronts of these waves have an appreciable curvature and they are not parallel. In the case of an actual grating with its many thousand lines ruled much closer together, the wave-fronts would be very nearly parallel. Draw AH and DK parallel to the axis of the telescope. The difference in the optical distances of H and K from the focal plane of the telescope is obviously equal to DG. DG = AD sin DAG = AD sin θ, where θ is the angle of diffraction of the rays DK and AH. Let e be the distance between two successive rulings. Then AD = $3e$ and DG = $3e \sin\theta$. The difference in the optical distances of H and K from the focal plane of the telescope is equal to three complete wave-lengths. Thus $3\lambda = 3e \sin\theta$, or $\lambda = e \sin\theta$, and the result obtained by considering the incident light as a single pulse agrees with the ordinary theory.

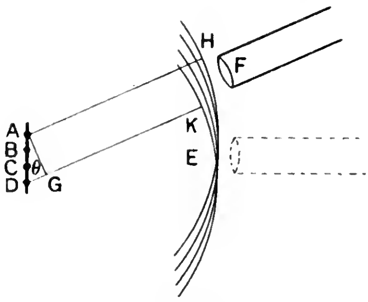


FIG. 291.

It is obvious that a single pulse will in this way produce a whole spectrum, whereas, if the incident light is regarded as consisting of a superposition of real cosine waves with different periods, we require a very great number of such waves to produce a continuous spectrum. Each cosine wave taken by itself produces only a single line. A single pulse may be produced by the collision of two atoms while the superposition of a great number of real cosine waves requires as source an elaborate vibrating system. Hence in comparing the two explanations the advantage of simplicity lies on the side of the pulse theory.

In explaining the second and third order grating spectra on the pulse theory we require to take into account the dispersion in the lenses of the instrument and in the air. In the case of the echelon we require to consider the dispersion in the glass.

The action of a grating on a light pulse is in some measure analogous to the action of the siren on an air-blast. The siren is an instrument which can change an irregular air-blast into a periodic air-blast of constant frequency. We assume, of course, that the disc is driven otherwise than by the air-blast itself. It thus makes a regular wave out of an irregular one. So does the grating.

Action of a Prism on a Light Pulse. Let the pulse be incident normally on the face OA of the prism AOB (fig. 292), and let there be a screen on the other face restricting the length OB that can be used. Let b be the greatest thickness of the prism used. After emergence from the prism the light is received by a lens, which makes it converge to a point. The diameter of the lens is so large that it receives all the rays that pass through the aperture OB; the latter causes the diffraction, not the rim of the lens. We shall ignore reflection and absorption losses both in the prism and the lens. It is assumed for simplicity that the refracting angle of the prism is 30°.

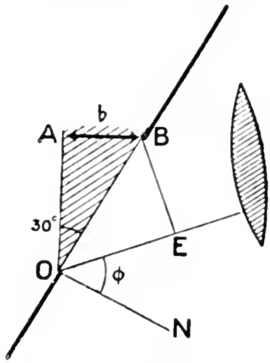


FIG. 292.

Let the incident light pulse be represented by

$$\int_0^\infty e^{-ha} \cos a(l - vt)\, a^{3/2}\, da,$$

i.e. it has the same analytical form as the depression of the surface in the hydrodynamical problem; l is distance from OA measured positive from left to right. The dispersion of the atmosphere is neglected; hence v is independent of a and the velocity of each of the components, into which the pulse can be resolved, remains the same as long as the pulse is in air.

The treatment depends on the principle of the superposition of small vibrations. Let us suppose we have a series of infinitely long trains of cosine waves incident on the prism and that the amplitude of each train is given in terms of its parameter by $e^{-ha} a^{3/2}$. These waves will be diffracted in all directions from the other face of the prism. Their phases after emergence are known since the index of refraction of the prism is known. Then, since the pulse is equivalent to the superimposed cosine trains before incidence, it must be equivalent to them after emergence.

The whole action of the prism depends on diffraction. Each point on the face OB must be considered as emitting rays of all possible wave-lengths in every direction. Take a definite direction OE, which is specified by the angle ϕ, which it makes with ON, the normal to the face OB. To this direction there corresponds a point P in the focal plane of the lens. Let the optical distance of P from O, measured along the ray, be s.

Let us consider the disturbance at P due to rays of parameter a. We have to sum an infinite number of rays of equal intensity and period but uniformly increasing phase. Let $2p$ be the difference of phase in radian measure between the first and last of them. Then, as is shown on p. 172, the amplitude of the resultant varies as

$$\frac{\sin p}{p}.$$

If BE is drawn perpendicular to OE and μ denotes the index of refraction of the prism, OE = OB sin ϕ = $2b \sin \phi$, since the angle of the prism is 30°, and consequently

$$2p = \frac{2\pi}{\lambda}(\mu b - 2b \sin \phi) = ab(\mu - 2 \sin \phi).$$

The phase of the resultant is the phase of the mean ray. Consequently the disturbance at P is given by

$$\frac{\sin p}{p} \cos \{a(s - vt) + p\}.$$

To get the whole disturbance at P we have to multiply the above expression by $e^{-ha}a^{3/2}$ in order to take account of the shape of the initial pulse and then integrate with respect to a. The result is

$$\int_0^\infty e^{-ha} \frac{\sin p}{p} \cos \{a(s - vt) + p\} a^{3/2} da \; ; \quad . \quad . \quad (4)$$

p is, of course, a function of a. For the expression to be integrable we are restricted to one function, $g - fa$. We shall assume this value. It will be shown further down that this assumption does not really restrict the generality of the results in any way. Meantime it should be noted that f and g are constants of the prism and independent of the shape of the incident pulse.

The variation of the factor $\frac{\sin p}{p}$ with p has already been studied on p. 173, and it has been shown that it has a maximum at $p = 0$, and decreases on both sides of this, becoming 0 at $p = \pm \pi$ and then oscillating about the axis, the amplitude of each successive oscillation being less than that of the one before it. Much the most important part of the curve is contained between the limits $p = \pm \pi$, i.e. between the limits $g - fa = \pm \pi$. Now we have

$$p = ab\left(\frac{\mu}{2} - \sin \phi\right) = g - fa,$$

hence

$$\frac{\mu}{2} = \sin \phi - \frac{f}{b} + \frac{g}{ba},$$

or

$$\frac{\mu}{2} = \sin \phi - \frac{f}{b} + \frac{g\lambda}{2\pi b}.$$

This gives

$$\frac{d\mu}{d\lambda} = \frac{g}{\pi b} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

In the case of crown glass in the neighbourhood of the D lines $d\mu/d\lambda$ has the value 500. Let us suppose that b has the value 2 cm. Then $g = 1000\,\pi$.

In the equation for the limits between which the principal variation of $\frac{\sin p}{p}$ takes place, namely $g - fa = \pm \pi$, g is 1000 times as great as π; hence while $\frac{\sin p}{p}$ varies from 0 to its maximum and back to 0 again, the value of a alters only by 0·2 per cent. Thus $e^{-ha}\, a^{3/2}$ practically remains stationary in the integral (4) while $\frac{\sin p}{p}$ varies, and can consequently be taken outside the integral sign. We have left then

$$\int_0^\infty \frac{\sin (g - fa)}{(g-fa)} \cos \{a(s - vt) + g - fa\}\, da.$$

Write $a - g/f = \beta$ and $g/f = a_1$. Then $\beta = a - a_1$ and the integral becomes

$$\int_{-a_1}^\infty \frac{\sin f\beta}{f\beta} \cos \{\beta(s - vt - f) + a_1(s - vt)\} d\beta$$

$$= \int_{-a_1}^\infty \frac{\sin \{\beta(s - vt) + a_1(s - vt)\} - \sin \{\beta(s - vt - 2f) + a_1(s - vt)\}}{2f\beta} d\beta$$

$$= \int_{-a_1}^\infty \frac{\sin \beta(s-vt) \cos a_1(s-vt) + \sin a_1(s-vt) \cos \beta(s-vt)}{2f\beta} - \ldots d\beta$$

$$= \cos a_1(s - vt) \int_{-a_1}^\infty \frac{\sin \beta(s - vt) - \sin \beta(s - vt - 2f)}{2f\beta} d\beta$$

$$+ \sin a_1(s - vt) \int_{-a_1}^\infty \frac{\cos \beta(s - vt) - \cos \beta(s - vt - 2f)}{2f\beta} d\beta$$

$$= \frac{1}{2f} \cos a_1(s - vt) \left[\int_{-a_1(s-vt)}^{\infty(s-vt)} \frac{\sin y}{y} dy - \int_{-a_1(s-vt-2f)}^{\infty(s-vt-2f)} \frac{\sin y}{y} dy \right]$$

$$+ \frac{1}{2f} \sin a_1(s - vt) \left[\int_{-a_1(s-vt)}^{\infty(s-vt)} \frac{\cos y}{y} dy - \int_{-a_1(s-vt-2f)}^{\infty(s-vt-2f)} \frac{\cos y}{y} dy \right].$$

The integrals in the brackets are known as the sine integral and cosine integral, and their values have been calculated for all values of y by J. W. L. Glaisher.*

The first bracket can be rewritten

$$\int_{-a_1(s-vt)}^{-a_1(s-vt-2f)} \frac{\sin y}{y} dy + \int_{\infty(s-vt-2f)}^{\infty(s-vt)} \frac{\sin y}{y} dy.$$

Let us consider the second term first. It will be shown farther down that $2f$ is positive and much smaller than s. When $t = 0$ both limits are $+\infty$; when $t = (s - 2)/v$ the lower limit changes to $-\infty$, and when $t = s/v$ the upper limit changes to $-\infty$. The value of the integral between the limits $\pm\infty$ is, of course, π. Hence from $t = (s - 2f)/v$ to $t = s/v$ the integral is π; before and after this it equals zero.

Next consider the other term. When $t = 0$ it is zero. As t approaches $(s - 2f)/v$ it suddenly approaches $\frac{1}{2}\pi$, then as the upper limit changes sign and becomes positive, it increases to π. It will be shown farther down that the upper limit practically reaches $+\infty$ long before the lower limit changes sign. As the lower limit becomes zero and then positive the integral decreases again to zero. The changes are the same as for the second term, only they take place continuously.

The second bracket can be written

$$\int_{-a_1(s-vt)}^{-a_1(s-vt-2f)} \frac{\cos y}{y} dy + \int_{\infty(s-vt-2f)}^{\infty(s-vt)} \frac{\cos y}{y} dy,$$

but this integral, though very great near the origin, diminishes rapidly to a small quantity, is zero at infinity, and so can be neglected in comparison with the other one.

* "Phil. Trans.," vol. 160, pp. 367-387 (1870).

Our final result, therefore, for the disturbance at P, when the term left outside the integral sign is included, is

$$\frac{\pi}{f} e^{-ha_1} a_1^{3/2} \cos a_1(s - vt)$$

between the times $t = (s - 2f)/v$ and $t = s/v$. Outside these limits there is no disturbance.

We had $g = 1000\pi$. If a_1 has the value for the D lines, $1{\cdot}066\ 10^5$, $f = g/a_1 = 0{\cdot}0293$ and is consequently small in comparison with s. Also $2fa_1 = 6{\cdot}28\ 10^3$. Hence as the sine integral differs only by 1 per cent from its value for infinity for $y = 75$, both assumptions made above are justified.

The number of complete wave-lengths in the train received at P is $\frac{2f}{\lambda} = \frac{a_1 f}{\pi} = \frac{g}{\pi} = b\frac{d\mu}{d\lambda}$, Rayleigh's expression for the resolving power. Its numerical value in this case is 1000.

The value assumed for p, namely, $g - fa$, leads to the result

$$\frac{d\mu}{d\lambda} = \frac{g}{\pi b},$$

i.e. that the dispersion is constant throughout the spectrum. We used the assumption, however, only for the short range during which $\frac{\sin p}{p}$ rose from 0 to its maximum value and fell back to zero again, and this range is so short, that for all the ordinary materials used for prisms the dispersion can be considered constant throughout it. Hence the generality of the result is in no way invalidated by the special expression assumed for p. It is obvious, too, that the proof is independent of the special angle assumed for the prism, and that it is not necessary for the pulse to be incident on the prism normally.

Also, if we determine that $g - fa$ is to represent p only within this range, it follows that

$$\frac{2f}{\lambda} = b\frac{d\mu}{d\lambda},$$

even if the dispersion is not constant throughout the spectrum, so that at every point in the spectrum and for all prism materials the number of waves in the train is equal to Rayleigh's expression for the resolving power.

In case some readers may have had difficulty in following all the above mathematics we shall sum up its results in words: If a single pulse is incident on the slit of a spectroscope, it gives rise to a continuous spectrum. If any one particular point of the spectrum is considered, the light arriving at that point consists of a train of cosine waves, the number of waves in the train being equal to Rayleigh's expression for the resolving power at that point in the spectrum.

It may be shown from Glaisher's tables that the train starts and stops quite suddenly. For example, in the case considered, the spectrum formed by a crown glass prism of 2 cm, base at a point near the D lines, where there are 1000 waves in the train, 97 per cent of the growth of the amplitude takes place within two wave-lengths, and the train ends equally suddenly.

Talbot's Bands. The fact that a succession of perfectly irregular impulses produces a continuous spectrum was established by Gouy in 1886 and independently by Rayleigh in 1889. It has also been developed by Schuster. Previous to that time the different monochromatic waves were held to have a real existence in the incident light. Rayleigh investigated the type of pulse that would have the same energy distribution as ordinary white light and found that the disturbance represented by

$$y = e^{-c^2x^2}$$

fulfilled the requirements to a considerable extent. This expression, however, neither represents the experimental results as well as the expression already used in this chapter, nor is it so convenient analytically.

The pulse theory of white light as compared with the old theory has, as already mentioned, the advantage that it requires simpler conditions in the source. There is also an experiment which seems to point definitely in its favour, namely Talbot's bands.

Talbot's bands are produced if, when viewing a continuous spectrum with a spectroscope, we insert a thin piece of glass such as a microscope cover glass between the eye and the eyepiece, so as to cover one half of the pupil. The bands are parallel to the Fraunhofer lines and are most distinct for a special thickness of the glass. Also it is essential that the glass be inserted from the side at which the blue of the spectrum appears. The glass may also be placed between the prism and telescope object glass instead of between the eyepiece and eye. In this case it must be inserted on the side of the thin edge of the prism.

The bands are clearly due to interference between the rays which pass through the thin glass plate and the rays which miss it, but on the old theory it is not easy to explain why they are visible only when the plate is inserted on the one side. The difficulty disappears if we regard the spectrum as produced from a single pulse.

According to the result proved on pp. 374-379 the pulse produces at every point in the spectrum a train containing $b\,d\mu/d\lambda$ wave-lengths. The first of these arrives at the time $t = s/v$, where s is measured from O and t from the instant at which the pulse is incident on OA; f is negative in all actual cases. Let us suppose that the screen at B is moved down to C. The effective value of b is halved and consequently the number of waves in the train is halved; but the first wave arrives at the same time as before. Hence the second half of the original train must be due to the light from the part BC.

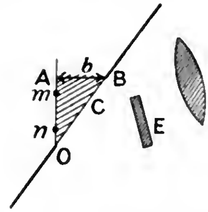


FIG. 293.

If a glass plate is inserted from the side E it retards the first half train, so that the two arrive superimposed and in a condition to interfere. If the retardation is $\frac{1}{2}b\ d\mu/d\lambda$ wave-lengths the superposition is perfect. Consequently in this case the interference bands are very black, and this gives the most favourable thickness of plate to insert. But if the glass plate is inserted from the other side it retards the second half of the train, a gap appears between the two halves, and interference is quite impossible.

The beams from the two halves of the prism face cross each other in the focal plane of the telescope. That is why, when the plate is inserted between eye and eyepiece, it must be put in from the other side.

The two trains interfere in the focal plane as if they came from m and n; if we move along the focal plane from a minimum to a maximum, the difference of the distances from m and n increases by $\lambda/2$. Hence the difference of the distances from A and O increases by λ. But when monochromatic light is used and the slit is made as narrow as possible, in moving from the central maximum to the first diffraction minimum on either side, the difference of the distances from A and O increases by λ (cf. p. 181). Thus the angular distance between two adjacent Talbot's bands is equal to twice the angular distance between the central maximum and first diffraction minimum when monochromatic light is used.

Monochromatic Radiation. So far we have been dealing with continuous spectra, such as the spectra of an electric glow lamp or of the crater of the arc, and it has been shown that in these cases the light before it enters the spectroscope consists probably of irregular pulses. How is it with monochromatic radiations such as, for example, the thallium green line?

In this case the wave is a periodic one before it enters the slit and can be represented approximately then by a sine or cosine curve. But there are differences. A sine wave goes on to infinity; it is limited neither in time nor space. But the thallium train falling on the spectroscope slit is a limited one. The amplitude of its oscillations increases from zero to a finite value, keeps near this value for a certain distance, and then decreases to zero again. Just as in the case of the single pulse, the thallium train can be regarded as due to the superposition of a number of fictitious cosine waves; only in the case of the thallium train the wave-lengths of these cosine waves vary only slightly from the wave-length of the thallium line, whereas in the case of the white light pulse they have all possible values.

Fizeau and Foucault used a Fresnel's mirror apparatus, illuminating the slit with sunlight, and placed the slit of a spectroscope to receive the light reflected by the two mirrors. The spectrum was crossed with dark bands parallel to the Fraunhofer lines showing that some of the colours were destroyed by interference. One of the mirrors was screwed forward parallel to itself; this caused the bands to move along the spectrum and to come closer together. The bands could still be detected in the neighbourhood of the F line when there was a difference of 1737 wave-lengths between the two paths; then they had come very close together in the spectrum. It was held formerly that this showed that white light consisted of regular trains, the vibrations of which took place without sensible change of phase for 1737 wave-lengths. On the pulse theory the two pulses which arrive by the different paths are each changed into a train by the spectroscope, and it is probable—a description of the spectroscope is not given—that these trains had distinctly more than 1737 waves in them. Hence even with this retardation they would overlap and interfere. The path difference is not a measure of the regularity of the components of the white light but a measure only of the resolving power of the spectroscope.

Quite otherwise is it with experiments in which there is no spectral resolution. Fizeau observed Newton's rings with an alcohol flame coloured with sodium as source of light and the usual method of convex surface on glass plate, only he had an arrangement by which the convex surface could be slowly raised above the plate. This made the rings move into the centre one after the other and disappear there. The rings were counted as they passed one particular point. Owing to the two sodium lines each producing its own system, the fringes became gradually fainter until, when 490 had passed, the bright rings of the one system fell on the dark rings of the other and the field had one uniform brightness. As the distance of the plates was increased, the fringes reappeared again and attained their former sharpness when 980 had passed. They disappeared again when 1470 had passed, and so on. Fizeau was able to count 52 groups of 980 rings before they disappeared finally, or about 50,000 altogether. This showed clearly that the sodium line trains from an alcohol flame contained 50,000 wave-lengths.

Similarly with his interferometer Michelson was able to obtain interference with the green line of mercury when there was a path difference of 540,000 wave-lengths. This shows that the source emits waves for 540,000 periods without a change of phase. A single prism was used to purify the light before it entered the interferometer, but its resolving power was insignificant compared to this number.

Action of a Prism on a Regular Train. We have already considered the action of a prism on a single pulse. It is interesting to consider its action on a regular train.

Suppose that the incident light is represented by

$$\cos \beta(x - vt),$$

but the wave starts suddenly, continues for N complete wave-lengths, and then stops suddenly. To find the effect of the prism on this train it is necessary first to reverse the ordinary procedure, and instead of analysing a pulse into waves, analyse the wave into pulses. Suppose the wave built up of a number of rectangular pulses, all of the same base length but having different heights. Then each of these pulses is dispersed by the prism into a system of wave trains; the wave trains for the different pulses differ among themselves in amplitude and phase. If we fix our attention on the point of the spectrum specified by a, in the limit when the pulses are made infinitely thin, the superposition of the wave trains gives at this point

$$\int \cos \beta v t_0 \cos a(x - v\{t - t_0\})\, dt_0 \quad . \quad . \quad . \quad (5)$$

$t = t_0$ gives the time at which the elementary pulse reaches the prism and $\cos \beta v t_0$ its height.

Let the resolving power at the wave-length in question be R, and let $2\pi R/a$, the length of the train into which a pulse is dispersed, be greater than $2\pi N/\beta$, the whole length of the incident train. Then for a length $2\pi(R/a - N/\beta)$ all the wave trains overlap completely; on each side of this there is a length $2\pi N/\beta$, in which the overlapping takes place echelonwise, at one end of which we have only a single train due to the first or last elementary pulse of the original train, and as we progress through which the trains due to the other elementary pulses are gradually added, until at the other end we have all the trains due to the initial wave. For the middle region the integral (5) is to be taken over a range $2\pi N/(\beta v)$; for the end regions, as we move outwards from the middle, the range gradually diminishes from this value to zero.

Let us suppose that $\beta \neq a$; then

$$\int \cos \beta v t_0 \cos a(x - v\{t - t_0\})dt_0 = \frac{1}{2(a+\beta)v} \sin\{a(x - vt) + vt_0(a+\beta)\}$$
$$+ \frac{1}{2(a - \beta)v} \sin\{a(x - vt) + vt_0(a - \beta)\}.$$

If $a = \beta$ the integral becomes

$$\frac{1}{4\beta v} \sin\{\beta(x - vt) + 2\beta v t_0\} + \frac{t_0}{2} \cos \beta(x - vt).$$

When the limits are substituted, it is obvious that it is only the second term of the second case that becomes important. Within the middle region it has a value

$$\frac{\pi N}{\beta v} \cos \beta(x - vt);$$

in the end regions the amplitude diminishes uniformly from $\pi N/(\beta v)$ to zero.

As was to be expected therefore the limited harmonic train has been refracted to one particular point in the spectrum. There it produces a train with N + R waves, during the first N of which the amplitude steadily increases, for the next R − N of which it is constant, and for the last N of which it decreases. In the train there are no irregular changes of phase.

But when $\beta \neq a$ the integral does not vanish. Consequently the other regions of the spectrum are not absolutely dark. With a bright line spectrum there must always be associated a faint background due simply to the fact that the harmonic trains producing the bright lines are limited. Theoretically it should be possible from a comparison of the intensities of the line and background to obtain a lower limit to the number of periods in the initial train.

Group Velocity. Let us suppose that two sine waves of equal amplitude but different wave-lengths λ_1 and λ_2 and different velocities v_1 and v_2 are superimposed on one another. Then the resultant wave is given by

$$A\left\{\sin\frac{2\pi}{\lambda_1}(x - v_1t) + \sin\frac{2\pi}{\lambda_2}(x - v_2t)\right\}$$

$$= 2A\sin\pi\left(x\left\{\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right\} - t\left\{\frac{v_1}{\lambda_1}+\frac{v_2}{\lambda_2}\right\}\right)\cos\pi\left(x\left\{\frac{1}{\lambda_1}-\frac{1}{\lambda_2}\right\} - t\left\{\frac{v_1}{\lambda_1}-\frac{v_2}{\lambda_2}\right\}\right).$$

Assume now that λ_1 is almost equal to λ_2, and v_1 almost equal to v_2; write $\lambda_1 = \lambda, v_1 = v, \lambda_2 = \lambda + d\lambda$ and $v_2 = v + dv$. Then, since $d\lambda$ and dv are small

$$\frac{1}{\lambda_1} - \frac{1}{\lambda_2} = \frac{1}{\lambda} - \frac{1}{\lambda + d\lambda} = \frac{d\lambda}{\lambda^2},$$

and
$$\frac{v_1}{\lambda_1} - \frac{v_2}{\lambda_2} = \frac{v}{\lambda} - \frac{v + dv}{\lambda + d\lambda} = \frac{vd\lambda - \lambda dv}{\lambda^2}.$$

On substituting these values in the expression for the wave we obtain

$$2A\sin\frac{2\pi}{\lambda}(x - vt)\cos\pi\left(x\frac{d\lambda}{\lambda^2} - t\frac{vd\lambda - \lambda dv}{\lambda^2}\right).$$

The wave-length of the cosine is, according to our assumption, much greater than the wave-length of the sine. Consequently the resultant wave is represented by something like the full curve in fig. 294.

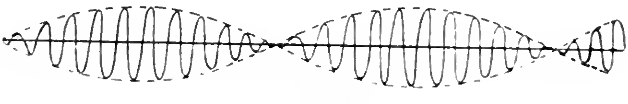


FIG. 294.

The dotted curve gives the cosine.

The individual waves move towards the right with a velocity v, the ordinary wave velocity. The dotted curve moves towards the right with a velocity given by

$$\frac{vd\lambda - \lambda dv}{d\lambda} = v - \lambda\frac{dv}{d\lambda} \quad . \quad . \quad . \quad . \quad (6)$$

This expression is called the group velocity. It is usually denoted by U.

If we return to the hydrodynamical problem studied on pp. 367-372 we find that the wave velocity was $\tau/(2\pi c)$. This gives

$$v^2 = \frac{v\tau}{2\pi c} = \frac{\lambda}{2\pi c}, \quad \text{or} \quad v = \sqrt{\frac{\lambda}{2\pi c}}.$$

Hence on applying (6) we obtain

$$\text{U} = \sqrt{\frac{\lambda}{2\pi c}} - \lambda\frac{d}{d\lambda}\sqrt{\frac{\lambda}{2\pi c}}$$
$$= \tfrac{1}{2}\sqrt{\frac{\lambda}{2\pi c}}.$$

The dominant wave-length was $\frac{8}{9}\pi h$. If we substitute this for λ we obtain

$$\text{U} = \tfrac{1}{2}\sqrt{\frac{1}{2\pi c}\,\frac{8\pi h}{9}} = \sqrt{\frac{h}{9c}},$$

which agrees with the group velocity as defined on p. 369. The definition here is wider. Every point in a group has its own group velocity. The group velocity is as much a characteristic of the wave-length and the medium as the wave velocity is. The wave velocity gives the rate of progression of the sine waves into which a group can be resolved; the group velocity gives the velocity of the point of reinforcement of two such waves of adjacent wave-length, i.e. the velocity of the characteristic features of the group.

The distinction between group velocity and wave velocity can be observed when a train of waves of approximately the same wave-length is advancing over the surface of deep water. The single waves are seen to advance through the group and die out as they approach the front, while their places are taken by fresh waves which appear at the rear.

Of course it is an exceptional case for groups to have the form represented in fig. 294, but the same formula can be derived for the general case by considering the velocity of the point of reinforcement of two neighbouring waves much in the same way as was done on p. 372.

In the case of interstellar space $dv/d\lambda = 0$. Hence it is v that is measured by Römer's method of determining the velocity of light. The terrestrial methods give U, but when the medium is air, the $dv/d\lambda$ term can be ignored. But in the case of Michelson's determination for carbon bisulphide by the rotating mirror method (p. 124), the correction has to be taken into account. He found the velocity in air 1·76 times the velocity in carbon bisulphide, whereas the index of refraction would give only the ratio 1·64. When the group velocity instead of the wave velocity in the carbon bisulphide is taken, the ratio comes out right.

EXAMPLES.

(1) Graph the expression

$$\frac{1}{x^{9/2}}\,e^{-t^2h/4cx^2}\cos\frac{t^2}{4cx}$$

for $t^2/c = 300$ and $h = \frac{1}{15}$.

(2) A single light pulse falls perpendicularly on a crown glass plate 1 cm. thick. How many waves of visible light will there be approximately in the group formed by its passage through the plate?

(3) Show that the pulse initially represented by

$$\frac{\cos \frac{5}{2} \tan^{-1} (x/h)}{(x^2 + h^2)^{5/4}}$$

has an energy distribution given by

$$\frac{c_1}{\lambda^5} e^{-c_2/\lambda T}$$

where c_1, c_2, and T are constants. (Wien's radiation formula, cf. Chapter XXV.)

(4) Show that the pulse initially represented by the series

$$\frac{\cos \frac{5}{2} \tan^{-1} (x/h)}{(x^2 + h^2)^{5/4}} + \frac{\cos \frac{5}{2} \tan^{-1} (x/2h)}{(x^2 + \{2h\}^2)^{5/4}} + \frac{\cos \frac{5}{2} \tan^{-1} (x/3h)}{(x^2 + \{3h\}^2)^{5/4}} \ldots$$

has an energy distribution given by

$$\frac{c_1}{\lambda^5} \left[e^{-c_2/\lambda T} + e^{-2c_2/\lambda T} + e^{-3c_2/\lambda T} \ldots \right]$$

or

$$\frac{c_1}{\lambda^5} \frac{1}{e^{c_2/\lambda T} - 1}$$

where c_1, c_2, and T are constants. (Planck's radiation formula, cf. Chapter XXV.)

CHAPTER XXII.

THE ELECTROMAGNETIC THEORY OF LIGHT.

THE electromagnetic theory of light is based on Clerk Maxwell's equations of the electromagnetic field. The latter can be stated as follows :—

$$\left.\begin{aligned} \frac{4\pi\sigma}{c}\mathsf{X} + \frac{\mathsf{K}}{c}\frac{\partial \mathsf{X}}{\partial t} &= \frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z} \\ \frac{4\pi\sigma}{c}\mathsf{Y} + \frac{\mathsf{K}}{c}\frac{\partial \mathsf{Y}}{\partial t} &= \frac{\partial\alpha}{\partial z} - \frac{\partial\gamma}{\partial x} \\ \frac{4\pi\sigma}{c}\mathsf{Z} + \frac{\mathsf{K}}{c}\frac{\partial \mathsf{Z}}{\partial t} &= \frac{\partial\beta}{\partial x} - \frac{\partial\alpha}{\partial y} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (7)$$

$$\left.\begin{aligned} \frac{\mu}{c}\frac{\partial\alpha}{\partial t} &= -\left(\frac{\partial \mathsf{Z}}{\partial y} - \frac{\partial \mathsf{Y}}{\partial z}\right) \\ \frac{\mu}{c}\frac{\partial\beta}{\partial t} &= -\left(\frac{\partial \mathsf{X}}{\partial z} - \frac{\partial \mathsf{Z}}{\partial x}\right) \\ \frac{\mu}{c}\frac{\partial\gamma}{\partial t} &= -\left(\frac{\partial \mathsf{Y}}{\partial x} - \frac{\partial \mathsf{X}}{\partial y}\right) \end{aligned}\right\} \quad . \quad . \quad . \quad (8)$$

σ, K, and μ are the electric conductivity, specific inductive capacity, and magnetic permeability of the medium at the point under consideration, c is the velocity of light in vacuo, X, Y, Z are the components of E, the electric intensity at the point, and α, β, γ are the components of H, the magnetic intensity at the point. σ, K, and X, Y, Z are measured in electrostatic units, and μ and α, β, γ in electromagnetic units.

The vector, the components of which are $\frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}, \frac{\partial\alpha}{\partial z} - \frac{\partial\gamma}{\partial x}, \frac{\partial\beta}{\partial x} - \frac{\partial\alpha}{\partial y}$, is said to be the curl of the vector whose components are α, β, γ. Hence the equations (7) and (8) can be contracted into

$$\frac{4\pi\sigma}{c}\mathsf{E} + \frac{\mathsf{K}}{c}\frac{\partial \mathsf{E}}{\partial t} = \text{curl } \mathsf{H}$$

and

$$\frac{\mu}{c}\frac{\partial \mathsf{H}}{\partial t} = -\text{ curl } \mathsf{E}.$$

Before proceeding to give an account of the proof of these equations it will be necessary to prove a theorem used in the proof.

Stokes' Theorem. The line integral of the tangential component of a vector taken round any closed curve is equal to the surface integral of the normal component of the curl of the same vector taken over any surface bounded by the curve, or

$$\int(\alpha dx + \beta dy + \gamma dz) = \iint\left\{l\left(\frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}\right) + m\left(\frac{\partial\alpha}{\partial z} - \frac{\partial\gamma}{\partial x}\right) + n\left(\frac{\partial\beta}{\partial x} - \frac{\partial\alpha}{\partial y}\right)\right\}dS.$$

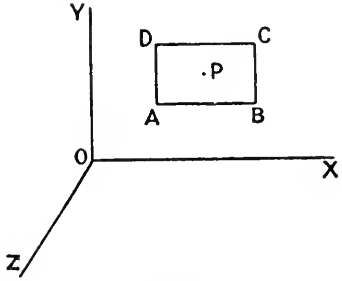


FIG. 295.

Let P (x, y, z) be the centre of a rectangle ABCD, the lengths of the sides of which are dx, dy, and let α, β, γ be the components of the vector at P.

At A and B α has respectively the values

$$\alpha - \frac{\partial\alpha}{\partial x}\frac{dx}{2} - \frac{\partial\alpha}{\partial y}\frac{dy}{2}, \quad \alpha + \frac{\partial\alpha}{\partial x}\frac{dx}{2} - \frac{\partial\alpha}{\partial y}\frac{dy}{2},$$

and at B and C β has respectively the values

$$\beta + \frac{\partial\beta}{\partial x}\frac{dx}{2} - \frac{\partial\beta}{\partial y}\frac{dy}{2}, \quad \beta + \frac{\partial\beta}{\partial x}\frac{dx}{2} + \frac{\partial\beta}{\partial y}\frac{dy}{2}.$$

Hence the average value of α on AB is $\alpha - \frac{\partial\alpha}{\partial y}\frac{dy}{2}$, and the average value of β on BC is $\beta + \frac{\partial\beta}{\partial x}\frac{dx}{2}$. Similarly the average value of α on DC is $\alpha + \frac{\partial\alpha}{\partial y}\frac{dy}{2}$, and the average value of β on AD is $\beta - \frac{\partial\beta}{\partial x}\frac{dx}{2}$. The line integral round the element is therefore

$$\left(\alpha - \frac{\partial\alpha}{\partial y}\frac{dy}{2}\right)dx + \left(\beta + \frac{\partial\beta}{\partial x}\frac{dx}{2}\right)dy - \left(\alpha + \frac{\partial\alpha}{\partial y}\frac{dy}{2}\right)dx - \left(\beta - \frac{\partial\beta}{\partial x}\frac{dx}{2}\right)dy$$

$$= \left(\frac{\partial\beta}{\partial x} - \frac{\partial\alpha}{\partial y}\right)dx\,dy.$$

Similar expressions hold when the rectangle is parallel to the yz or zx planes. In this way we infer that the equation holds for any small area having its planes parallel to a co-ordinate plane.

Now consider the triangular element ABC the normal to which is given by l, m, n. Since the contributions from OA, OB, OC cut out, the line integral round ABC is obviously equal to the sum of the line integrals round BCO, CAO, and ABO, that is to

$$\left(\frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}\right)\triangle\,\text{BCO} + \left(\frac{\partial\alpha}{\partial z} - \frac{\partial\gamma}{\partial x}\right)\triangle\,\text{CAO} + \left(\frac{\partial\beta}{\partial x} - \frac{\partial\alpha}{\partial y}\right)\triangle\,\text{ABO}$$

by the result already proved. This becomes

$$\left\{l\left(\frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}\right) + m\left(\frac{\partial\alpha}{\partial z} - \frac{\partial\gamma}{\partial x}\right) + n\left(\frac{\partial\beta}{\partial x} - \frac{\partial\alpha}{\partial y}\right)\right\}\Delta$$

where Δ is the area of ABC. Hence we infer that the equation holds for any small area whatever.

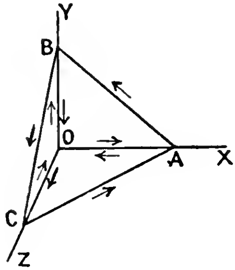


FIG. 296.

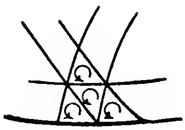


FIG. 297.

Take any surface and divide it up into elementary triangles. The line integral round the surface is equal to the sum of the line integrals round the individual triangles, because, as may be seen from fig. 297, every side of a triangle not at the same time on the bounding edge of the surface is traversed twice during the integration in different directions, and so contributes nothing to the total. Hence the line integral round the surface is equal to

$$\iint\left\{l\left(\frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}\right) + m\left(\frac{\partial\alpha}{\partial z} - \frac{\partial\gamma}{\partial x}\right) + n\left(\frac{\partial\beta}{\partial x} - \frac{\partial\alpha}{\partial y}\right)\right\}dS,$$

which proves the theorem.

§ The above proof of Stokes' theorem, although satisfactory from the physical standpoint, has nevertheless two gaps in it, somewhat irritating to a mathematician. Hence the following proof may be substituted; it is more logical though physically less instructive.

$$\iint\left(m\frac{\partial\alpha}{\partial z} - n\frac{\partial\alpha}{\partial y}\right)dS = \iint\frac{\partial\alpha}{\partial z}m\,dS - \iint\frac{\partial\alpha}{\partial y}n\,dS = \iint\frac{\partial\alpha}{\partial z}dz\,dx - \iint\frac{\partial\alpha}{\partial y}dx\,dy$$
$$= \int(\alpha_3 - \alpha_2)\,dx - \int(\alpha_1 - \alpha_2)\,dx = \int(\alpha_3 - \alpha_1)\,dx.$$

As may be seen from a diagram, α_1 and α_3 are the values of α for the points in which a plane parallel to YZ cuts the curve, and α_2 is the value for the point in which the lines through these two points intersect. Now

$$\int(\alpha_3 - \alpha_1)\,dx = \int\alpha\,dx$$

when regard is paid to the sign of dx at the two points on the curve. By treating the other terms in the surface integral in the same way the theorem follows.

Derivation of the First Three Equations. When an electric current flows along a wire it produces a magnetic field in the neighbourhood of that wire. This was discovered by Oersted in 1820; he found, when a straight wire was held horizontal in the meridian above a magnet which was free to turn about a vertical axis, that the magnet was deflected when a current passed along the wire. There was a force on the north pole due to the current tending to move it in a circle round the wire in one direction, and there was a force on the south pole tending to move it round the wire in the other direction. It has been found as a result of experiment, that if it were possible for the force due to the current to carry the north pole right round the wire to its original position, then the work done by the current on the pole would be $4\pi im/c$, where i is the strength of the current in electrostatic units and m is the strength of the pole. If, on the contrary, the pole were moved round the current against the magnetic field of the latter, then this quantity of work would be done against the current. The result holds true no matter how irregular the path pursued by the pole is. It is necessary only that the latter should be brought back to its starting-point and that it should go round the wire only once. The wire carrying the current does not require to be straight; indeed it is not necessary for the current to be carried by a wire at all. Suppose we have a mass of metal through every point of which a current is flowing and we draw in imagination a circuit inside this mass, then if a pole of strength m were carried round this imaginary circuit, the work done would be given by the same theorem. This theorem is known as the first circuital theorem. There will, of course, be a definite magnetic intensity at every point inside the mass of metal.

Let us put the first circuital theorem into a more general form. Let all space be filled with a conducting medium, not necessarily homogeneous, and let there be electric currents everywhere. At the point x, y, z let the components of H be a, β, γ, and let the components of current per unit area be u, v, w. That is, if we set up an area of 1 sq. cm. at right angles to the x axis, u gives the quantity of electricity measured in electrostatic units which flows through it in one second.

Draw any closed circuit in this medium. Then, by Stokes' theorem,

$$\int(adx + \beta dy + \gamma dz) = \iint\left\{l\left(\frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}\right) + m\left(\frac{\partial a}{\partial z} - \frac{\partial\gamma}{\partial x}\right) + n\left(\frac{\partial\beta}{\partial x} - \frac{\partial a}{\partial y}\right)\right\}d\mathsf{S} \quad . \quad . \quad . \quad . \quad (9)$$

The expression on the right of (9) is the surface integral of the normal component of the curl of H taken over any surface bounded by the circuit. The expression on the left is the line integral of the tangential component of H taken round the circuit. This may easily be seen, for, if we consider the element ds, its direction cosines are dx/ds, dy/ds, and dz/ds; similarly the direction cosines of H are a/H, β/H, and γ/H. The tangential component of H multiplied by the length of the element is

$$\mathsf{H} \cos\theta\, ds,$$

where θ is the angle between H and ds. But

$$\cos\theta = \frac{a}{\mathsf{H}}\frac{dx}{ds} + \frac{\beta}{\mathsf{H}}\frac{dy}{ds} + \frac{\gamma}{\mathsf{H}}\frac{dz}{ds}.$$

Hence

$$\mathsf{H}\cos\theta\, ds = adx + \beta dy + \gamma dz.$$

By the first circuital theorem

$$\int(\alpha dx + \beta dy + \gamma dz) = \frac{4\pi}{c}\iint(lu + mv + nw)dS.$$

The surface integral on the right gives the total current through the circuit. Combining this equation with (9) we obtain

$$\frac{4\pi}{c}\iint(lu + mv + nw)dS = \iint\left\{l\left(\frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}\right) + m\left(\frac{\partial\alpha}{\partial z} - \frac{\partial\gamma}{\partial x}\right) + n\left(\frac{\partial\beta}{\partial x} - \frac{\partial\alpha}{\partial y}\right)\right\}dS.$$

The above equation is true, no matter what the boundaries and shape of the surface are. It holds true for every element of it, no matter what values l, m, n may have; we may therefore equate the two integrands. Thus the equation decomposes into the following three:

$$\frac{4\pi u}{c} = \frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}, \quad \frac{4\pi v}{c} = \frac{\partial\alpha}{\partial z} - \frac{\partial\gamma}{\partial x}, \quad \frac{4\pi w}{c} = \frac{\partial\beta}{\partial x} - \frac{\partial\alpha}{\partial y}, \quad . \quad (10)$$

which hold for every point in the medium. They are equations which enable us to determine the current when the magnetic intensity is known.

The Displacement Current. The first circuital theorem was stated originally for steady currents. Now, according to the modern theory, an electric current in a wire consists of a procession of electrons along the wire. Electrons are small particles which have all the same mass m and carry the same negative charge e. Fig. 298 represents some-

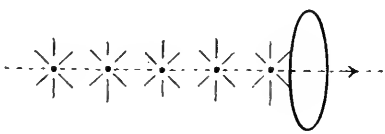


FIG. 298.

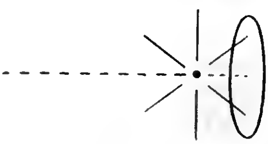


FIG. 299.

what crudely a procession of electrons going through a circuit. The lines of force are shown diverging from each electron. If N electrons pass through the circuit per second, the current strength is $-Ne$ and the work done in carrying unit pole round the circuit is numerically equal to $4\pi Ne/c$.

The first circuital theorem was stated at first only for the case of currents in wires, when a large number of electrons passes through the circuit per second. What happens when the electron itself does not pass through the circuit but only a portion of its field does? The second diagram above will make this question clear. Suppose that a single electron comes from the left along the dotted line, reaches the position shown in fig. 299 but stops there and goes no farther. It does not go through the circuit itself but part of its field does. Will there be a magnetic intensity round the circuit?

From an electron there issue $4\pi e$ lines of electric intensity or lines of force.* When the electron is at a distance from the plane of the

* The lines of force represent the electric intensity both in direction and magnitude. Their direction gives the direction of the intensity, and the number

circuit, the number of its lines of force through the circuit is zero. As it approaches the plane of the circuit this number increases, and, when it reaches the plane, the number is $2\pi e$. As it travels to infinity on the other side, the number increases to $4\pi e$, so that altogether there is a change of $4\pi e$ lines. When N electrons pass through the circuit per second there is a change of $4\pi\mathsf{N}e$ lines per second. The work done in carrying unit magnetic pole round the circuit is in this case $4\pi\mathsf{N}e/c$, i.e. it is equal to the rate of change in the number of lines divided by c. We can therefore regard the magnetic field as due to the rate of change in the number of the lines instead of to the passage of the electrostatic charges. Both methods of regarding the phenomenon lead to the same result when we are dealing with conduction currents in wires.

But in the case represented in fig. 299 they lead to different results. For in this case there is a change in the number of lines through the circuit but no charge goes through it. Which method leads to the correct result in this case?

It is the fundamental feature of Clerk Maxwell's theory that he regarded the magnetic field as due to the change in the number of lines per second.* If there is an increase in the number of lines through a circuit due to certain electrons, while the electrons themselves do not go through the circuit, then, according to his theory, there is a *displacement* current through the circuit measured by the rate of change of the number of lines of force divided by 4π, i.e. measured by $1/(4\pi)$ times the rate of change of the surface integral of electric intensity taken over the surface bounded by the circuit. The resultant current, the current which is to be substituted in equations (10), is the sum of the displacement and conduction currents.

If we have a point charge e situated in a medium of specific inductive capacity K, the electric intensity E at a point distant r from it is given by

$$\mathsf{E} = \frac{e}{\mathsf{K}r^2}.$$

The number of lines of electric intensity issuing from the charge in this medium is $4\pi e/\mathsf{K}$. If N point charges pass per second through a circuit in this medium, the work done in carrying unit magnetic pole round the circuit is still $4\pi\mathsf{N}e/c$. But the rate of change of the number of lines of intensity is in this case only $4\pi\mathsf{N}e/\mathsf{K}$. Hence the work done in this case is K/c times the rate of change of the number of lines of

intersecting unit of area gives the component of the intensity, in the direction at right angles to the area.

* This is not Maxwell's way of arriving at the displacement current, but students find it easier and it is certainly more in accordance with modern thought. Maxwell regarded the conduction current as continuous and different in kind from the displacement current, while here there is really no distinction. It is merely an accident that the charge goes through the circuit in the one case. The theory of displacement currents was given by Maxwell in 1865, while the idea of regarding the conduction current as due to the motion of electrons dates only from 1898.

intensity, and the displacement current through a circuit is $K/(4\pi)$ times the rate of change of the surface integral of electric intensity taken over the surface bounded by the circuit.

The conduction current per unit area in the direction of Ox is equal to σX. This may be seen by considering a cube of unit side; $1/\sigma$ is the resistance of this cube, and X is numerically equal to the electromotive force between two opposite faces. To obtain the displacement current per unit area in the direction of Ox consider a face of this cube perpendicular to the x axis. The number of lines of intensity through it is X. Consequently the required displacement current per unit area is

$$\frac{K}{4\pi}\frac{\partial X}{\partial t}.$$

If we substitute the sum of these expressions for u in equations (10) and at the same time make the corresponding substitutions for v and w we obtain

$$\frac{4\pi\sigma}{c}X + \frac{K}{c}\frac{\partial X}{\partial t} = \frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z},$$
$$\frac{4\pi\sigma}{c}Y + \frac{K}{c}\frac{\partial Y}{\partial t} = \frac{\partial\alpha}{\partial z} - \frac{\partial\gamma}{\partial x},$$
$$\frac{4\pi\sigma}{c}Z + \frac{K}{c}\frac{\partial Z}{\partial t} = \frac{\partial\beta}{\partial x} - \frac{\partial\alpha}{\partial y},$$

the first three equations of the electromagnetic field.

Example on the First Circuital Theorem. A point charge of electricity is situated in air on the axis of a circle of radius b at a distance a from the plane of the circle. Let P (fig. 300) be the position of the point charge, and let AB be the trace of the circle. With P as centre draw the sphere ADB. Then the area of the cap ADB is equal to the area intercepted on the enveloping cylinder by the planes AB and DF, i.e. 2π DF × CD. CD = PD − PC = PA − PC = $(a^2 + b^2)^{\frac{1}{2}} - a$. Thus the area of the cap is

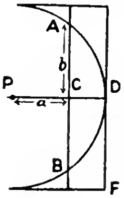


FIG. 300.

$$2\pi(a^2 + b^2)^{\frac{1}{2}}\{(a^2 + b^2)^{\frac{1}{2}} - a\}.$$

The electric intensity at D is

$$\frac{e}{a^2 + b^2}.$$

Hence the displacement current through the circle is

$$\frac{1}{4\pi}\frac{d}{dt}\left\{2\pi(a^2 + b^2)^{\frac{1}{2}}\{(a^2 + b^2)^{\frac{1}{2}} - a\}\frac{e}{a^2 + b^2}\right\}$$
$$= \frac{e}{2}\frac{d}{dt}\left\{1 - \frac{a}{(a^2 + b^2)^{\frac{1}{2}}}\right\}.$$
$$= \frac{e}{2}\left\{-\frac{\frac{da}{dt}}{(a^2 + b^2)^{\frac{1}{2}}} + \frac{a^2\frac{da}{dt}}{(a^2 + b^2)^{3/2}}\right\}$$
$$= -\frac{e}{2}\frac{da}{dt}\frac{b^2}{(a^2 + b^2)^{3/2}}.$$

Let H denote the tangential component of the magnetic field round the circle. Then, by the first circuital theorem,

$$2\pi b\,\mathsf{H} = -\frac{4\pi}{c}\frac{e}{2}\frac{da}{dt}\frac{b^2}{(a^2+b^2)^{3/2}}$$

or

$$\mathsf{H} = -\frac{eb}{c(a^2+b^2)^{3/2}}\frac{da}{dt}.$$

da/dt is the velocity of the point charge away from the circle. If we write $\mathsf{AP} = r$, $\angle\mathsf{APC} = \theta$ and $-da/dt = v$, then this result may be written

$$\mathsf{H} = \frac{ev\sin\theta}{cr^2}.$$

Thus the magnetic intensity increases from zero as the point charge approaches; it reaches a maximum when the charge passes, and then finally diminishes to zero again.

Derivation of the Second Three Equations of the Electromagnetic Field. If a coil of wire is connected in circuit with a galvanometer and the pole of a magnet is thrust into the coil, a current is set up through the galvanometer. This current endures as long as the magnet is moving, and ceases whenever the magnet comes to rest. If, again, instead of thrusting the magnet into the coil a current is started or stopped in a neighbouring circuit, a transient current is set up in the first circuit. This transient current lasts only as long as the value of the current in the neighbouring circuit is altering, and ceases whenever the latter attains a steady value. Such currents are called induced currents and their laws were determined experimentally by Faraday.

In both the above cases the magnitude of the induced current depends on the resistance of the circuit, i.e. on the material of which the wire is composed. The induced electromotive force, that is, the resistance of the circuit multiplied by the current, is the same no matter what the material of the circuit is. We shall now proceed to give the mathematical expression for the induced electromotive force.

Suppose we take any surface bounded by the circuit and divide this surface up into elements of area which are so small that they can be regarded as plane. Let $d\mathsf{S}$ be the area of one of these elements, let α, β, γ be the components of the magnetic intensity on its surface, and let l, m, n be the direction cosines of its normal. Then $l\alpha + m\beta + n\gamma$ is the component of the magnetic intensity normal to the surface,

$$(l\alpha + m\beta + n\gamma)d\mathsf{S}$$

gives the number of lines of magnetic intensity through the element, and

$$\iint(l\alpha + m\beta + n\gamma)d\mathsf{S}$$

gives the number of lines of magnetic intensity through the whole circuit.

Then if the circuit is situated in air, $-1/c$ times the time rate of change of the above integral gives the induced electromotive force round it. If the circuit is situated in an isotropic medium, not air, then the induced electromotive force round it is given by

$$-\frac{1}{c}\frac{\partial}{\partial t}\iint\mu\,(l\alpha + m\beta + n\gamma)d\mathsf{S}$$

where μ is a constant termed the magnetic permeability. This theorem, which embodies all Faraday's results, is sometimes called the second circuital equation. If the medium is not homogeneous μ may vary from place to place, i.e. it is a function of S. For air μ is unity.*

If in a certain region we have a changing magnetic field, brought about either by the motion of a magnetic pole or by making and breaking a current in a neighbouring circuit, and if we place a circuit anywhere in this region, there is always an induced current and consequently an induced electromotive force round the circuit. The electromotive force is the line integral of the electric intensity round the circuit. Now the circuit may be placed in an infinite number of positions; consequently we are led to the conclusion that as a result of the changing magnetic field there is an electric intensity with a definite magnitude and direction at every point in the region.

Let ds be an element of length of the circuit and let θ be the angle which it makes with the direction of E, then the second circuital equation can be written

$$c\int\mathsf{E}\cos\theta\,ds = -\frac{\partial}{\partial t}\iint\mu\,(l\alpha + m\beta + n\gamma)d\mathsf{S}\,. \qquad . \quad (11)$$

The minus sign means that if the surface integral of normal induction is increasing, the direction of the increase is connected with the line integral of electric intensity in the manner typified by a left-handed screw. By Stokes' theorem

$$c\int\mathsf{E}\cos\theta\,ds = c\iint\left\{l\left(\frac{\partial\mathsf{Z}}{\partial y} - \frac{\partial\mathsf{Y}}{\partial z}\right) + m\left(\frac{\partial\mathsf{X}}{\partial z} - \frac{\partial\mathsf{Z}}{\partial x}\right) + n\left(\frac{\partial\mathsf{Y}}{\partial x} - \frac{\partial\mathsf{X}}{\partial y}\right)\right\}d\mathsf{S} \quad . \ (12)$$

On combining (11) and (12) we obtain

$$\frac{\partial}{\partial t}\iint\mu(l\alpha + m\beta + n\gamma)d\mathsf{S} = -c\iint\left\{l\left(\frac{\partial\mathsf{Z}}{\partial y} - \frac{\partial\mathsf{Y}}{\partial z}\right) + m\left(\frac{\partial\mathsf{X}}{\partial z} - \frac{\partial\mathsf{Z}}{\partial x}\right) + n\left(\frac{\partial\mathsf{Y}}{\partial x} - \frac{\partial\mathsf{X}}{\partial y}\right)\right\}d\mathsf{S}.$$

This equation is true no matter what the boundaries and shape of the surface are. It is consequently true for every element of the surface, no matter what the values of l, m, n may be. It therefore decomposes into the following three equations,

$$\frac{\mu}{c}\frac{\partial\alpha}{\partial t} = -\left(\frac{\partial\mathsf{Z}}{\partial y} - \frac{\partial\mathsf{Y}}{\partial z}\right),\quad \frac{\mu}{c}\frac{\partial\beta}{\partial t} = -\left(\frac{\partial\mathsf{X}}{\partial z} - \frac{\partial\mathsf{Z}}{\partial x}\right),\quad \frac{\mu}{c}\frac{\partial\gamma}{\partial t} = -\left(\frac{\partial\mathsf{Y}}{\partial x} - \frac{\partial\mathsf{X}}{\partial y}\right),$$

Maxwell's second three equations for the electromagnetic field.

Equation for an Electromagnetic Wave. Let σ, the conductivity of the medium, be zero. Then the equations of the electromagnetic field become

* $\iint\mu(l\alpha + m\beta + n\gamma)\,d\mathsf{S}$ is said to give the magnetic induction through the circuit.

$$\frac{K}{c}\frac{\partial X}{\partial t} = \frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z}, \quad \frac{K}{c}\frac{\partial Y}{\partial t} = \frac{\partial \alpha}{\partial z} - \frac{\partial \gamma}{\partial x}, \quad \frac{K}{c}\frac{\partial Z}{\partial t} = \frac{\partial \beta}{\partial x} - \frac{\partial \alpha}{\partial y},$$

$$-\frac{\mu}{c}\frac{\partial \alpha}{\partial t} = \frac{\partial Z}{\partial y} - \frac{\partial Y}{\partial z}, \quad -\frac{\mu}{c}\frac{\partial \beta}{\partial t} = \frac{\partial X}{\partial z} - \frac{\partial Z}{\partial x}, \text{ and } -\frac{\mu}{c}\frac{\partial \gamma}{\partial t} = \frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}.$$

Differentiating the first with regard to t and substituting from the last two we obtain

$$\frac{K}{c}\frac{\partial^2 X}{\partial t^2} = \frac{\partial^2 \gamma}{\partial y \partial t} - \frac{\partial^2 \beta}{\partial z \partial t} = -\frac{c}{\mu}\left(\frac{\partial^2 Y}{\partial x \partial y} - \frac{\partial^2 X}{\partial y^2} - \frac{\partial^2 X}{\partial z^2} + \frac{\partial^2 Z}{\partial x \partial z}\right),$$

i.e. $$\frac{\mu K}{c^2}\frac{\partial^2 X}{\partial t^2} = \frac{\partial^2 X}{\partial x^2} + \frac{\partial^2 X}{\partial y^2} + \frac{\partial^2 X}{\partial z^2} - \frac{\partial}{\partial x}\left(\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z}\right) \quad . \quad . \quad (13)$$

Let us assume that there are no electrostatic charges in the region under consideration. Then

$$\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z} = 0 \quad . \quad . \quad . \quad . \quad . \quad (14)$$

For, if we suppose a single electrostatic charge of magnitude e_1, at the point x_1, y_1, z_1 the intensity due to it at a point P (x, y, z) distant r_1 from x_1, y_1, z_1, is given by e_1/r_1^2. Taking the components of this parallel to the axes we obtain,

$$X_1 = \frac{e_1}{r_1^2}\frac{x - x_1}{r_1}, \quad Y_1 = \frac{e_1}{r_1^2}\frac{y - y_1}{r_1}, \text{ and } Z_1 = \frac{e_1}{r_1^2}\frac{z - z_1}{r_1}.$$

This gives $$\frac{\partial X_1}{\partial x} = \frac{\partial}{\partial x}\frac{e_1(x - x_1)}{r_1^3} = \frac{e_1}{r_1^3} - \frac{3e_1(x - x_1)}{r_1^4}\frac{\partial r_1}{\partial x}.$$

But $r_1^2 = (x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2$ and hence $2r_1\frac{\partial r_1}{\partial x} = 2(x-x_1)$

or $$\frac{\partial r_1}{\partial x} = \frac{x - x_1}{r_1}.$$

Substitute this value in the above equation and we obtain

$$\frac{\partial X_1}{\partial x} = \frac{e_1}{r_1^3} - \frac{3e_1(x - x_1)}{r_1^4}\frac{(x - x_1)}{r_1} = \frac{e_1}{r_1^3}\left(1 - \frac{3(x - x_1)^2}{r_1^2}\right).$$

Similarly $$\frac{\partial Y_1}{\partial y} = \frac{e_1}{r_1^3}\left(1 - \frac{3(y - y_1)^2}{r_1^2}\right) \text{ and } \frac{\partial Z_1}{\partial z} = \frac{e_1}{r_1^3}\left(1 - \frac{3(z - z_1)^2}{r_1^2}\right).$$

Hence $$\frac{\partial X_1}{\partial x} + \frac{\partial Y_1}{\partial y} + \frac{\partial Z_1}{\partial z} = \frac{e_1}{r_1^3}\left(3 - \frac{3\{(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2\}}{r_1^2}\right) = 0.$$

Similarly, if we assume a charge e_2 at a point x_2, y_2, z_2, and consider its effect at x, y, z, we find that

$$\frac{\partial X_2}{\partial x} + \frac{\partial Y_2}{\partial y} + \frac{\partial Z_2}{\partial z} = 0,$$

and by assuming a number of such charges we find finally that

$$\frac{\partial}{\partial x}(X_1 + X_2 \ldots) + \frac{\partial}{\partial y}(Y_1 + Y_2 \ldots) + \frac{\partial}{\partial z}(Z_1 + Z_2 \ldots) = 0.$$

It is assumed, of course, that none of the charges is at the point x, y, z.

If we substitute (14) in (13), then the latter equation reduces to

$$\frac{\mu K}{c^2}\frac{\partial^2 X}{\partial t^2} = \frac{\partial^2 X}{\partial x^2} + \frac{\partial^2 X}{\partial y^2} + \frac{\partial^2 X}{\partial z^2} \quad . \quad . \quad . \quad . \quad (15)$$

The Equation of Wave Motion. Equation (15) is the same as

$$\frac{\partial^2 X}{\partial t^2} = v^2\left(\frac{\partial^2 X}{\partial x^2} + \frac{\partial^2 X}{\partial y^2} + \frac{\partial^2 X}{\partial z^2}\right), \quad . \quad . \quad . \quad . \quad (16)$$

an equation which occurs frequently in physics. The solutions of this equation represent waves travelling with velocity v.

This may be seen by considering some special cases. Suppose, for example, that X does not vary with y and z. Then the equation reduces to

$$\frac{\partial^2 X}{\partial t^2} = v^2 \frac{\partial^2 X}{\partial x^2}. \quad . \quad . \quad . \quad . \quad . \quad (17)$$

In order to solve this equation write

$$x_1 = x - vt, \quad x_2 = x + vt.$$

Then
$$\frac{\partial}{\partial x} = \frac{\partial}{\partial x_1}\frac{\partial x_1}{\partial x} + \frac{\partial}{\partial x_2}\frac{\partial x_2}{\partial x} = \frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2}.$$

Therefore
$$\frac{\partial^2 X}{\partial x^2} = \frac{\partial^2 X}{\partial x_1^2} + 2\frac{\partial^2 X}{\partial x_1 \partial x_2} + \frac{\partial^2 X}{\partial x_2^2}.$$

Also
$$\frac{\partial}{\partial t} = \frac{\partial}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial}{\partial x_2}\frac{\partial x_2}{\partial t} = -v\frac{\partial}{\partial x_1} + v\frac{\partial}{\partial x_2},$$

and consequently
$$\frac{\partial^2 X}{\partial t^2} = v^2\frac{\partial^2 X}{\partial x_1^2} - 2v^2\frac{\partial^2 X}{\partial x_1 \partial x_2} + v^2\frac{\partial^2 X}{\partial x_2^2}.$$

On substituting these values, the original equation reduces to

$$\frac{\partial^2 X}{\partial x_1 \partial x_2} = 0.$$

The most general solution of this is obviously

$$X = f_1(x_1) + f_2(x_2).$$

Hence the most general solution of the original equation is

$$X = f_1(x - vt) + f_2(x + vt).$$

Now, if $f_1(x - vt)$ be plotted as a function of x, it is exactly the same as $f_1(x)$ in shape, but every point on it is displaced a distance vt to the right of the corresponding point in $f_1(x)$. It thus represents an irregular wave travelling towards the right with uniform velocity v, the shape of the wave at time $t = 0$ being given by $X = f_1(x)$. Similarly $X = f_2(x + vt)$ represents a wave travelling towards the left with uniform velocity v. The general solution is the sum of these two waves.

If we write

$$f_1(x - vt) = \sin\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) \quad \text{and} \quad f_2(x + vt) = \sin\frac{2\pi}{\tau}\left(t + \frac{x}{v}\right),$$

then the two irregular waves reduce to infinite trains of harmonic waves such as have been used earlier in this book.

Suppose the wave is plane but goes in a direction inclined to the co-ordinate axes. Let l, m, n be the cosines of the angles which its direction makes with the co-ordinate axes. Then, as may be verified by substitution, the solution of the equation takes the form $f(lx + my + nz - vt)$ in the case of an irregular wave, and $\sin\frac{2\pi}{\tau}\left(t - \frac{lx + my + nz}{v}\right)$ in the case of an infinite harmonic one.

§ Let us suppose that X varies only with r, its distance from the origin. Then

$$\frac{\partial X}{\partial x} = \frac{\partial X}{\partial r}\frac{\partial r}{\partial x}.$$

But $r^2 = x^2 + y^2 + z^2$ and $2r\frac{\partial r}{\partial x} = 2x$; hence

$$\frac{\partial X}{\partial x} = \frac{\partial X}{\partial r}\frac{x}{r}.$$

Further
$$\frac{\partial^2 X}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial X}{\partial r}\frac{x}{r}\right) = \frac{\partial X}{\partial r}\frac{1}{r} + x\frac{\partial}{\partial r}\left(\frac{\partial X}{\partial r}\frac{1}{r}\right)\frac{x}{r}$$
$$= \frac{\partial X}{\partial r}\frac{1}{r} + \frac{x^2}{r}\left(\frac{\partial^2 X}{\partial r^2}\frac{1}{r} - \frac{1}{r^2}\frac{\partial X}{\partial r}\right) = \frac{x^2}{r^2}\frac{\partial^2 X}{\partial r^2} + \frac{\partial X}{\partial r}\left(\frac{1}{r} - \frac{x^2}{r^3}\right).$$

Similarly
$$\frac{\partial^2 X}{\partial y^2} = \frac{y^2}{r^2}\frac{\partial^2 X}{\partial r^2} + \frac{\partial X}{\partial r}\left(\frac{1}{r} - \frac{y^2}{r^3}\right)$$

and
$$\frac{\partial^2 X}{\partial z^2} = \frac{z^2}{r^2}\frac{\partial^2 X}{\partial r^2} + \frac{\partial X}{\partial r}\left(\frac{1}{r} - \frac{z^2}{r^3}\right).$$

On substituting these values in (16) we obtain

$$\frac{\partial^2 X}{\partial t^2} = v^2\left\{\frac{\partial^2 X}{\partial r^2}\Sigma\frac{x^2}{r^2} + \frac{\partial X}{\partial r}\left(\frac{3}{r} - \Sigma\frac{x^2}{r^3}\right)\right\}$$
$$= v^2\left(\frac{\partial^2 X}{\partial r^2} + \frac{2}{r}\frac{\partial X}{\partial r}\right) = \frac{v^2}{r}\frac{\partial^2}{\partial r^2}(rX),$$

which may be written
$$\frac{\partial^2}{\partial t^2}(rX) = v^2\frac{\partial^2}{\partial r^2}(rX) \quad . \quad . \quad . \quad . \quad (18)$$

This is the same as (17) except for the fact that rX takes the place of X and r takes the place of x. The most general solution of (18) consequently takes the form

$$rX = f_1(r - vt) + f_2(r + vt),$$

or
$$X = \frac{1}{r}f_1(r - vt) + \frac{1}{r}f_2(r + vt) \quad . \quad . \quad (19)$$

This obviously represents two spherical waves of irregular form both travelling with uniform velocity v, the first away from the origin and the second towards the origin. The amplitude of each wave is inversely proportional to its distance from the origin.

A particular form of (19) is

$$\frac{A}{r}\sin\frac{2\pi}{T}\left(t - \frac{r}{v}\right).$$

This represents an infinite harmonic spherical wave travelling away from the origin.

Electromagnetic Waves. Equation (15) shows that X is propagated by wave motion, the velocity of the waves being $c/\sqrt{\mu K}$. We can prove the same for Y, Z, α, β, γ by proceeding in exactly the same way. Consequently the electric and magnetic intensities are propagated in a dielectric with a velocity $c/\sqrt{\mu K}$, and are said to constitute an electromagnetic wave. Now for air $\mu = 1$ and $K = 1$ when they are measured respectively in electromagnetic and electrostatic units, as is done here. Hence for air the velocity is equal to c.

It is shown in books on electricity, that, when the same electrical quantity is measured on both the electrostatic and electromagnetic systems of units, the ratio of the results is always a power of c. For example the numerical value of the capacity of a condenser on the electrostatic system is c^2 times its value on the electromagnetic system. The value on the electrostatic system can be found by calculation, and the value on the electromagnetic system by charging the condenser to a known potential and then discharging it through a galvanometer. Thus the value of c can be determined by purely electrical methods. The following table, taken from Kaye and Laby's "Physical and Chemical Constants," gives the results of the most important determinations, together with the names of the observers :—

c.	Observer.	c.	Observer.	c.	Observer.
$\times 10^{10}$		$\times 10^{10}$		$\times 10^{10}$	
2·963	J. J. Thomson, 1883	2·997	Thomson and Searle, 1890	3·001	Hurmuzescu, 1896
2·982	Rowland, 1889	3·009	Pellat, 1891	2·997	Perot and Fabry
3·000	Rosa, 1889	2·993	Abraham, 1892	2·997	Rosa and Dorsey, 1907

The mean of all these determinations is $2·993\ 10^{10}$ in cms./sec. The most modern determinations of the velocity of light in air by both the rotating mirror and toothed wheel methods gave $2·9986\ 10^{10}$ cms./sec.

The velocity of light in air is thus the same as that of an electromagnetic wave in air. This striking result was first published by Maxwell in 1865. It leads at once to the conclusion that light itself is an electromagnetic wave, but this conclusion was not accepted until Hertz performed his experiments in 1887-88, the reason for the delay being that Maxwell's reasoning was purely mathematical and did not attract attention; besides, before Hertz's work there were no experiments which could be explained only by Maxwell's theory of electromagnetism and not by the previous theories.

Hertz's Experiments. According to the theory that prevailed before Maxwell electric and magnetic intensity were propagated with infinite velocity. If we have a point charge of electricity e at a point A (fig. 301) the electric intensity at P is e/AP^2. Now suppose the point charge to move instantaneously from A to B. According to the old theory of action at a distance the magnitude of the intensity at P changed instantaneously to e/BP^2. According to Maxwell's theory the change takes some time to travel out to P. While the charge moves from A to B it is equivalent to an electric current and is, therefore, surrounded by a magnetic field. A magnetic intensity is propagated out to P but it lasts only as long as the electric intensity is changing. It is the propagation of this change in the electric intensity together with the attendant magnetic intensity, that constitutes an electromagnetic wave. If we were to move a charged pith ball suddenly from one point to another by mechanical means, although a wave would be produced, the effect would not be nearly powerful enough to observe. Hertz produced the waves by means of a piece of apparatus called a vibrator or oscillator.

One form of vibrator consisted of two rectangular brass plates, each of about 40 sq. cms. area, to each of which was attached a thick wire with a spherical brass knob at its end. By means of the wires shown in the diagram these knobs were connected to the terminals of the secondary of an induction coil. The two plates act like the plates of a condenser of low capacity. When the primary of the induction coil is broken, a high electromotive force is induced in the secondary. This causes a current to flow out of the one plate, through the wire, round the secondary of the induction coil, and through the wire into the other plate. The two plates thus acquire charges of different sign and the condenser becomes charged. This goes on until the potential difference between the two knobs is too great for the insulation of the

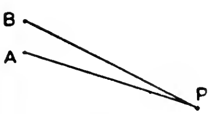


FIG. 301.

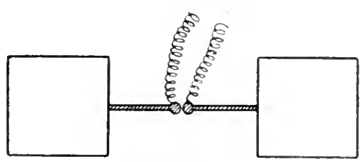


FIG. 302.

air gap between them to withstand. It therefore breaks down and a discharge passes across the gap. This discharge is an oscillatory one. The current first goes the one way until the two plates acquire charges of the opposite sign to those which they possessed at first. It then reverses until the plates are charged the same way as at first. It then reverses again, and so on, the electricity surging backwards and forwards across the gap until its energy is dissipated and the gap ceases to conduct. Every time the primary of the induction coil is broken, the whole process is repeated over again. The number of complete vibrations in one discharge is not large, about four, and their period is not quite constant, as the resistance of the gap is changing all the time. Instead of the pith ball moving from A to B in fig. 301 we have in the Hertzian oscillator shown in fig. 302 an electric charge moving backwards and forwards between two plates with a complete period of about $1{\cdot}4\ 10^{-8}$ sec. This is, of course, very much faster than any charged conductor could be moved by mechanical means. For the oscillator in fig. 302 to work properly it is necessary for the surfaces of the knobs to be polished very smooth so that the discharge starts suddenly.

To detect the waves produced by the oscillator Hertz used a resonator consisting of a piece of wire bent into a circle, the ends terminating in brass knobs the length of the air gap between which could be regulated with a micrometer screw. The diameter of the circle was chosen so that the period of the oscillations which could take place in the wire coincided with the period of the oscillator itself.

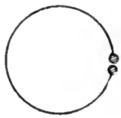


FIG. 303.

If the resonator was placed so that owing to the action of the oscillator the magnetic induction through it changed, a periodic electromotive force was induced round the circle. This caused an oscillating difference of potential between the two knobs, the amplitude of which increased until finally sparks passed between them. The maximum length of the gap across which sparks would pass was a measure of the strength of the induction through the circle, and consequently a measure of the component of magnetic intensity at right angles to the plane of the circle.

If we call all the planes through the axis of the oscillator in fig. 302 meridian planes, and the field of the oscillator be explored with a resonator, then it is found, that at a distance from the oscillator the direction of the electric intensity at any point is in the meridian plane and at right angles to the line joining the point to the oscillator, while the direction of the magnetic intensity is at right angles to the meridian plane. Since the disturbance is travelling out from the oscillator, it follows that the electric and magnetic intensities are at right angles to one another and to their direction of propagation.

The most convincing way of showing that there actually are waves travelling out from the oscillator is to produce interference. To do this they are allowed to fall perpendicularly on a metal screen 2 or 3

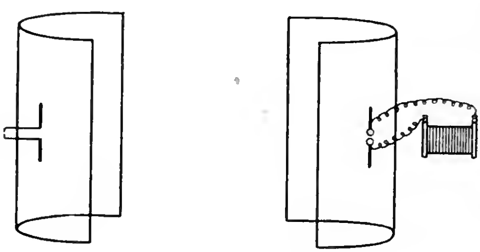


FIG. 304.

metres square. This reflects them and makes them interfere with the incident waves, and the result is that stationary waves are formed in front of the screen. The resonator sparks at the loops and does not spark at the nodes.

Hertz also demonstrated the refraction of electromagnetic waves by a prism. For this purpose he made a vibrator consisting of two brass cylinders, which were placed in the focal line of a parabolic zinc reflector. Thus after reflection the waves proceeded as a parallel beam. The resonator consisted of two pieces of thick wire placed in the focal line of a similar reflector; the spark gap was taken through to the back of the mirror so that it could be observed without obstructing the rays. When the beam fell on the reflector, it was concentrated on the resonator and sparks were produced in the gap. When a large prism of pitch of refracting angle 30° was placed in the path of the beam, the sparks stopped but were produced again when the receiving mirror was moved a distance to the side, showing clearly that the beam was deviated by the prism. The angle of minimum deviation was 22° so that the index of refraction of the prism for the electromagnetic waves worked out at 1·69.

If the receiving mirror is rotated round the direction of the beam coming from the vibrator, the sparks diminish in length, and when the directions of the resonator and vibrator are at right angles to one another the sparks cease altogether. This is because the original beam is plane-polarised; its electric intensity is parallel to the focal line of the first mirror, and, when the mirrors are crossed, its component in the direction of the resonator is zero. This plane polarisation can also be shown experimentally in an interesting way by means of a wooden frame, on which a number of wires have been wound with their directions parallel, so as to form a kind of diffraction grating. If this is placed between the vibrator and receiver with the wires parallel to the former, it is opaque to the radiation. But when the direction of the wires is at right angles to the axis of the vibrator, the waves are transmitted freely. The action of the wire screen is thus analogous to that of a Nicol's prism.

Effect of the Medium on the Velocity. The general expression for the velocity of an electromagnetic wave is $c/\sqrt{\mu K}$. For all transparent bodies μ differs inappreciably from 1; hence the velocity reduces to $c/\sqrt{K}$. But the velocity of light in a medium of index of refraction n is c/n. Hence $n = \sqrt{K}$, or, in other words, the index of refraction is equal to the square root of the specific inductive capacity.

If we proceed to test this relation we find that for certain gases it is approximately true. For example, we have the following results:—

	n.	$\sqrt{K}$.
Air	1·000294	1·000295
Hydrogen	1·000138	1·000132
Carbon dioxide	1·000449	1·000473

The values of n are for the middle of the spectrum. But for most substances there is a wide difference. Thus for water $n = 1·33$ and $\sqrt{K} = 9$. The difference is due to the fact that indices of refraction are determined for light waves, i.e. for rapidly changing fields, while specific inductive capacities are determined in the laboratory for stationary fields, i.e. for waves of an infinite length. Thus agreement is to be expected only when the index of refraction does not change with the wave-length.

Propagation of a Plane Wave. Consider the expression

$$Y = B \cos \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right).$$

It represents a plane wave of electric intensity propagated in the direction of the positive x axis, v being the velocity of the wave, τ its period, and B the maximum value of its amplitude. In any plane parallel to yz the electric intensity at any time has everywhere the same value. If we fix our attention on a definite plane, then, as time progresses, the electric intensity undergoes a simple harmonic variation. If we fix our attention on a definite time and move the plane instantaneously in the direction of the x axis, then the electric intensity again undergoes a simple harmonic variation when regarded as a function of the distance. Its direction, however, always remains parallel to the y axis.

Put X = Z = 0, and substitute for X, Y, and Z in the second three equations of the electromagnetic field. Then

$$-\frac{\mu}{c}\frac{\partial \alpha}{\partial t} = 0, \quad -\frac{\mu}{c}\frac{\partial \beta}{\partial t} = 0, \quad -\frac{\mu}{c}\frac{\partial \gamma}{\partial t} = B\frac{2\pi}{\tau v}\sin\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right).$$

The constants of integration must be zero, as there are supposed to be no permanent magnets or steady currents in the field. We thus obtain

$$\alpha = \beta = 0, \quad \gamma = B\frac{c}{\mu v}\cos\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right).$$

This represents a plane wave of magnetic intensity, of the same period and velocity as the former wave propagated in the same direction, the magnetic intensity in the wave being always parallel to the Z axis. According to the equations of the electromagnetic field, we cannot have the one wave without the other. Both together are said to constitute a plane electromagnetic wave plane polarised in the zx plane. In the wave the electric and magnetic intensities are at right angles both to one another and to the direction of propagation.

Apart altogether from the differential equation we can think of the propagation of the above waves occurring as follows: Let us imagine that initially there is an electric intensity in the y direction everywhere in a thin layer parallel to the plane $x = 0$. This induces a magnetic intensity in the z direction on the front surface of the layer and an equal one in the $-z$ direction on the back surface of the layer. The first magnetic intensity induces an electric intensity in the $+y$ direction in front of the layer and one in the $-y$ direction inside the layer. This latter cancels out the initial electric intensity. In this way a pulse is propagated.

Similarly, if we had started out with the wave

$$Z = C\cos\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right),$$

by substitution in the equations of the electromagnetic field we would have found associated with it the wave

$$\beta = -\mathrm{C}\frac{c}{\mu v}\cos\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right).$$

Together they constitute a plane electromagnetic wave plane polarised in the xy plane.

§ Suppose that the plane polarised wave is not propagated in the direction of one of the coordinate axes but in any direction whatever, the direction cosines of which are l, m, n. Then it may be represented by

$$\mathrm{X} = \mathrm{A}\cos\frac{2\pi}{\tau}\left(t - \frac{lx + my + nz}{v}\right), \quad \mathrm{Y} = \mathrm{B}\cos\frac{2\pi}{\tau}\left(t - \frac{lx + my + nz}{v}\right),$$

$$\mathrm{Z} = \mathrm{C}\cos\frac{2\pi}{\tau}\left(t - \frac{lx + my + nz}{v}\right).$$

Now

$$\frac{\partial \mathrm{X}}{\partial x} + \frac{\partial \mathrm{Y}}{\partial y} + \frac{\partial \mathrm{Z}}{\partial z} = 0.$$

Substituting in this equation and cancelling out the common factor

$$\frac{2\pi}{\tau v}\sin\frac{2}{\tau}\left(t - \frac{lx + my + nz}{v}\right)$$

we obtain $l\mathrm{A} + m\mathrm{B} + n\mathrm{C} = 0$, i.e. A, B, and C are not independent, but the resultant electric intensity must be at right angles to the direction of propagation.

Substituting in the second three equations of the electromagnetic field, we obtain

$$-\frac{\mu}{c}\frac{\partial a}{\partial t} = (m\mathrm{C} - n\mathrm{B})\frac{2\pi}{\tau v}\sin\frac{2\pi}{\tau}\left(t - \frac{lx + my + nz}{v}\right),$$

$$-\frac{\mu}{c}\frac{\partial \beta}{\partial t} = (n\mathrm{A} - l\mathrm{C})\frac{2\pi}{\tau v}\sin\frac{2\pi}{\tau}\left(t - \frac{lx + my + nz}{v}\right),$$

$$-\frac{\mu}{c}\frac{\partial \gamma}{\partial t} = (l\mathrm{B} - m\mathrm{A})\frac{2\pi}{\tau v}\sin\frac{2\pi}{\tau}\left(t - \frac{lx + my + nz}{v}\right),$$

whence

$$a = (m\mathrm{C} - n\mathrm{B})\frac{c}{\mu v}\cos\frac{2\pi}{\tau}\left(t - \frac{lx + my + nz}{v}\right),$$

$$\beta = (n\mathrm{A} - l\mathrm{C})\frac{c}{\mu v}\cos\frac{2\pi}{\tau}\left(t - \frac{lx + my + nz}{v}\right),$$

$$\gamma = (l\mathrm{B} - m\mathrm{A})\frac{c}{\mu v}\cos\frac{2\pi}{\tau}\left(t - \frac{lx + my + nz}{v}\right).$$

We see from the form of the coefficients that (a, β, γ) is at right angles to both (X, Y, Z) and (l, m, n).

Poynting's Theorem. If a conductor receives an electrostatic charge, the energy of the charge is stored up in the field. This can be shown very well with a Leyden jar, the inner and outer coatings of which can be detached from the glass. If the condenser is insulated and charged, and if it is taken apart with insulating tongs and the two coatings put into contact with one another, no spark passes between them. But if it is put together again and then discharged, the spark is as great as it would have been had the condenser never been taken apart. The energy of the charge has apparently been stored up in the glass.

The energy of a system of charged conductors can, of course, be calculated from the work done in bringing each elementary charge from infinity, in analogy with the method of calculating the potential energy of a system of gravitating masses. It is found that the same numerical value can always be obtained by assuming that there is an amount of electrostatic energy stored at every point of the field equal to

$$\frac{\mathsf{KE}^2}{8\pi}$$

per unit volume, E being the electric intensity at the point.

Similarly, at a point in a magnetic field, where H is the magnetic intensity, we assume that there is a quantity of energy $\frac{\mu\mathsf{H}^2}{8\pi}$ per unit volume. This assumption gives the same value for the energy of a system of electric circuits as is obtained by using the equivalence of each circuit to a magnetic shell.

We assume, therefore, that the density of the total energy in the field is given by

$$\frac{1}{8\pi}(\mathsf{KE}^2 + \mu\mathsf{H}^2).$$

Suppose now that we have a certain region of space bounded by a closed surface. The energy in this region is given by

$$\iiint\frac{1}{8\pi}(\mathsf{KE}^2 + \mu\mathsf{H}^2)dx\,dy\,dz$$

$$= \iiint\frac{1}{8\pi}\{\mathsf{K}(\mathsf{X}^2 + \mathsf{Y}^2 + \mathsf{Z}^2) + \mu(a^2 + \beta^2 + \gamma^2)\}dx\,dy\,dz,$$

the integration being taken throughout the whole region. The rate of increase of the energy in the region is obtained by differentiating the integral with respect to t, and is equal to

$$\iiint\frac{1}{4\pi}\left\{\mathsf{K}\left(\mathsf{X}\frac{\partial\mathsf{X}}{\partial t} + \mathsf{Y}\frac{\partial\mathsf{Y}}{\partial t} + \mathsf{Z}\frac{\partial\mathsf{Z}}{\partial t}\right) + \mu\left(a\frac{\partial a}{\partial t} + \beta\frac{\partial\beta}{\partial t} + \gamma\frac{\partial\gamma}{\partial t}\right)\right\}dx\,dy\,dz.$$

Substituting for $\mathsf{K}\frac{\partial\mathsf{X}}{\partial t}, \ldots, \ldots, \mu\frac{\partial a}{\partial t}, \ldots, \ldots$, from the equations of the electromagnetic field, this becomes

$$\frac{c}{4\pi}\iiint\left\{\Sigma\mathsf{X}\left(\frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}\right) - \Sigma a\left(\frac{\partial\mathsf{Z}}{\partial y} - \frac{\partial\mathsf{Y}}{\partial z}\right)\right\}dx\,dy\,dz$$

$$= \frac{c}{4\pi}\iiint\left\{\frac{\partial}{\partial x}(\beta\mathsf{Z} - \gamma\mathsf{Y}) + \frac{\partial}{\partial y}(\gamma\mathsf{X} - a\mathsf{Z}) + \frac{\partial}{\partial z}(a\mathsf{Y} - \beta\mathsf{X})\right\}dx\,dy\,dz$$

$$= \frac{c}{4\pi}\iint\{l(\beta\mathsf{Z} - \gamma\mathsf{Y}) + m(\gamma\mathsf{X} - a\mathsf{Z}) + n(a\mathsf{Y} - \beta\mathsf{X})\}d\mathsf{S}$$

by Gauss's theorem. The vector the components of which are $\beta\mathsf{Z} - \gamma\mathsf{Y}$, $\gamma\mathsf{X} - a\mathsf{Z}$, and $a\mathsf{Y} - \beta\mathsf{X}$, is evidently at right angles to both H and E, and its numerical value is equal to

$$\{(\beta\mathsf{Z} - \gamma\mathsf{Y})^2 + (\gamma\mathsf{X} - a\mathsf{Z})^2 + (a\mathsf{Y} - \beta\mathsf{X})^2\}^{\frac{1}{2}} = \mathsf{EH}\sin\theta,$$

where θ is the angle between the directions of H and E.

The surface integral is the surface integral of the normal component of $\frac{c\mathsf{EH}\sin\theta}{4\pi}$ taken over the surface bounding the region. It is natural then to interpret $\frac{c\mathsf{EH}\sin\theta}{4\pi}$ at a point in space as the rate of flow of energy per unit area at that point. This result is due to Prof. Poynting.

Energy of a Plane Wave. Suppose that the wave is propagated in the x direction and that it is polarised in the zx plane. Then it may be represented by

$$\mathsf{Y} = \mathsf{B}\cos\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right), \quad \gamma = \mathsf{B}\frac{c}{\mu v}\cos\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right).$$

By Poynting's theorem the energy is flowing in the direction of the x axis, i.e. the direction of the flow of energy is identified with the ray, and the rate of flow at any time for any value of x is given by

$$\frac{c\mathsf{EH}\sin\theta}{4\pi} = \frac{c^2}{4\pi\mu v}\mathsf{B}^2\cos^2\frac{2\pi}{\tau}\left(t - \frac{x}{v}\right) \text{ ergs/sq. cm. sec.}$$

This expression oscillates between zero and a constant positive value, but never changes sign. The energy flow is therefore always forward. The period of the oscillations is so small that they cannot be detected by the eye or any physical instrument; it is the mean value that is important. Now the mean value of $\cos^2\theta$, between $\theta = 0$ and $\theta = \pi$, is $\frac{1}{2}$. Hence the intensity of the wave is equal to

$$\frac{c^2}{8\pi\mu v}\mathsf{B}^2.$$

As all our observations on light are made in air, for all practical purposes we may put $\mu v = c$. The intensity of the wave is therefore proportional to the square of the amplitude, a result which has already (cf. p. 131) been deduced from other considerations.

Boundary Conditions. It is now necessary to determine the conditions that must be fulfilled at the boundary of two media when an electromagnetic wave passes from the one to the other. To fix our ideas, let the xy plane be the boundary, let the specific inductive capacity of the upper medium be K, of the lower medium K', and take the axis of z positive downwards. We shall also suppose that as we pass through the boundary, the specific inductive capacity changes discontinuously from the value K to K'.

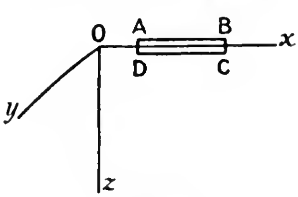


FIG. 305.

Consider the rectangle ABCD, the side AB of which is in the one medium and the side CD in the other, both AB and CD being extremely close to Ox. Let a unit magnetic pole be carried round this rectangle. Then the work done against the field must be zero, because the area of the rectangle is so small that the displacement current flowing through it may be neglected. The work done on the ends AD and BC may be neglected owing to their being so small. Thus the work done on AB must be equal and opposite to the work done on CD, or, in other words, the magnetic intensities along AB and DC are equal. We arrive, therefore, at the condition that the tangential component of magnetic intensity must have the same magnitude and direction on both sides of the boundary, that is, in this case α and β must be the same on both sides of the boundary.

Similarly, by taking unit positive electric charge round the rectangle, it may be shown that X and Y have the same values on both sides of the boundary.

EXAMPLES.

(1) Show that, when an electromagnetic wave passes from one medium to another, if the normal to the surface of separation is parallel to the z axis, then KZ and $\mu\gamma$ have the same values on both sides of the surface of separation.

(2) A plane wave, which is plane polarised in the xy plane, is propagated in the x direction. Find an expression for the energy in one wave-length, given that the magnetic intensity is specified by

$$\beta = \mathrm{A} \sin \frac{2\pi}{\tau}\left(t - \frac{x}{v}\right).$$

(3) A long straight cylindrical wire of circular section is carrying a steady current C, measured in electrostatic units. Calculate by Poynting's theorem the rate of flow of energy into a length d of the wire, given that R is the resistance of this length, and compare the result with Joule's expression for the heat evolved in this portion of the wire.

(4) Assuming that $\mu = 1$ but that K has the value K_1, K_2, or K_3 according as it is associated with X, Y, or Z, so that the first three equations of the electromagnetic field become

$$\frac{\mathrm{K}_1}{c}\frac{\partial \mathrm{X}}{\partial t} = \frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z}, \quad \frac{\mathrm{K}_2}{c}\frac{\partial \mathrm{Y}}{\partial t} = \frac{\partial \alpha}{\partial z} - \frac{\partial \gamma}{\partial x}, \quad \frac{\mathrm{K}_3}{c}\frac{\partial \mathrm{Z}}{\partial t} = \frac{\partial \beta}{\partial x} - \frac{\partial \alpha}{\partial y},$$

show that, if a plane wave is propagated, the resultant electric intensity is not at right angles to the direction of propagation but the resultant magnetic intensity is. Also show that the velocity in the direction l, m, n is given by Fresnel's law for the velocity of light in a crystal (cf. p. 205), namely

$$\frac{l^2}{\mathrm{A}^2 - v^2} + \frac{m^2}{\mathrm{B}^2 - v^2} + \frac{n^2}{\mathrm{C}^2 - v^2} = 0,$$

where $\mathrm{A}^2 = c^2/\mathrm{K}_1$, $\mathrm{B}^2 = c^2/\mathrm{K}_2$, and $\mathrm{C}^2 = c^2/\mathrm{K}_3$.

CHAPTER XXIII.

REFLECTION AND REFRACTION.

The General Case. Let a plane polarised plane wave of monochromatic light fall upon the plane boundary of two transparent media. Take the axis of Z positive downwards, and let the boundary of the two media be given by $z = 0$. Let the plane of incidence be the zx plane and let the angle of incidence be ϕ. Let the specific inductive capacity of the upper medium be K and of the lower medium K′. As is usual in problems in optics, we put the magnetic permeability of both media equal to unity.

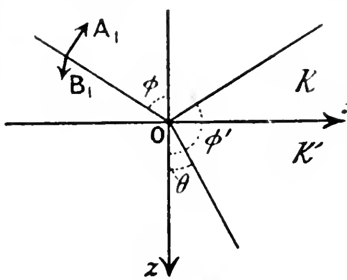


FIG. 306.

Resolve the electric intensity in the incident wave into two components, of maximum amplitude A_1 in the plane of incidence and B_1 perpendicular to the plane of incidence. Then the plane of polarisation of the incident light makes an angle $\cot^{-1} B_1/A_1$ with the zx plane.

Resolve A_1 into components $A_1 \cos\phi$ parallel to Ox and $-A_1 \sin\phi$ parallel to Oz. Then the electric intensity in the incident wave may be written

$$X_1 = A_1 \cos\phi \cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{K}\{x\sin\phi + z\cos\phi\}}{c}\right),$$

$$Y_1 = B_1 \cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{K}\{x\sin\phi + z\cos\phi\}}{c}\right),$$

$$Z_1 = -A_1 \sin\phi\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{K}\{x\sin\phi + z\cos\phi\}}{c}\right),$$

since the velocity in air is c and the direction cosines of the normal to the wave-front are $\sin\phi$, 0, $\cos\phi$.

To find the magnetic intensity associated with the electric intensity, substitute for X_1, Y_1, Z_1 in the second three equations of the electromagnetic field and solve for $\alpha_1, \beta_1, \gamma_1$, making the constants of integration zero. Then we obtain

$$\alpha_1 = -\mathsf{B}_1\sqrt{\mathsf{K}}\cos\phi\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}}\{x\sin\phi + z\cos\phi\}}{c}\right),$$

$$\beta_1 = +\mathsf{A}_1\sqrt{\mathsf{K}}\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}}\{x\sin\phi + z\cos\phi\}}{c}\right),$$

$$\gamma_1 = +\mathsf{B}_1\sqrt{\mathsf{K}}\sin\phi\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}}\{x\sin\phi + z\cos\phi\}}{c}\right).$$

The above six equations represent the whole incident wave. When it arrives at the boundary it gives rise to a refracted and a reflected wave. We shall assume that the maximum values of the electric intensity of the refracted wave are respectively A_2 for the component in, and B_2 for the component perpendicular to the plane of incidence. We then obtain the following equations for the refracted wave simply by substituting for A_1, B_1, K and ϕ:

$$\mathsf{X}_2 = \mathsf{A}_2\cos\theta\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}'}\{x\sin\theta + z\cos\theta\}}{c}\right),$$

$$\mathsf{Y}_2 = \mathsf{B}_2\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}'}\{x\sin\theta + z\cos\theta\}}{c}\right),$$

$$\mathsf{Z}_2 = -\mathsf{A}_2\sin\theta\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}'}\{x\sin\theta + z\cos\theta\}}{c}\right),$$

$$\alpha_2 = -\mathsf{B}_2\sqrt{\mathsf{K}'}\cos\theta\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}'}\{x\sin\theta + z\cos\theta\}}{c}\right),$$

$$\beta_2 = +\mathsf{A}_2\sqrt{\mathsf{K}'}\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}'}\{x\sin\theta + z\cos\theta\}}{c}\right),$$

$$\gamma_2 = +\mathsf{B}_2\sqrt{\mathsf{K}'}\sin\theta\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}'}\{x\sin\theta + z\cos\theta\}}{c}\right).$$

Similarly, for the reflected wave, we obtain

$$\mathsf{X}_3 = \mathsf{A}_3\cos\phi'\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}}\{x\sin\phi' + z\cos\phi'\}}{c}\right),$$

$$\mathsf{Y}_3 = \mathsf{B}_3\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}}\{x\sin\phi' + z\cos\phi'\}}{c}\right),$$

$$\mathsf{Z}_3 = -\mathsf{A}_3\sin\phi'\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}}\{x\sin\phi' + z\cos\phi'\}}{c}\right),$$

$$\alpha_3 = -\mathsf{B}_3\sqrt{\mathsf{K}}\cos\phi'\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}}\{x\sin\phi' + z\cos\phi'\}}{c}\right),$$

$$\beta_3 = +\mathsf{A}_3\sqrt{\mathsf{K}}\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}}\{x\sin\phi' + z\cos\phi'\}}{c}\right),$$

$$\gamma_3 = +\mathsf{B}_3\sqrt{\mathsf{K}}\sin\phi'\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{\mathsf{K}}\{x\sin\phi' + z\cos\phi'\}}{c}\right).$$

In the above A_3 and B_3 are put respectively equal to the components of the maximum electric intensity in and perpendicular to the plane of incidence. Also, we do not assume that ϕ', the angle of reflection, is equal to ϕ, the angle of incidence. It should be noted that ϕ' is the angle that the normal to the reflected wave-front makes with the *positive* direction of Oz.

We have now to apply the boundary conditions. A_1, B_1, ϕ are known and we wish to determine A_2, B_2, θ, A_3, B_3, ϕ'. It is at once clear that for $z = 0$ all the components of electric and magnetic intensity must be proportional to the same function of x and t, i.e.

$$\sqrt{K} \sin \phi = \sqrt{K'} \sin \theta = \sqrt{K} \sin \phi'.$$

This equation contains the laws of refraction and reflection, for it may be written

$$\frac{\sin \phi}{\sin \theta} = \sqrt{\frac{K'}{K}} = n, \quad \phi = \pi - \phi'.$$

The laws of reflection and refraction are thus derivable from the mere fact that there are boundary equations, and they do not depend on the particular form of the latter.

Since the tangential components of the electric and magnetic intensities are the same on both sides of the boundary, we have $X_1 + X_3 = X_2$ with three similar equations. These give

$$(A_1 - A_3) \cos \phi = A_2 \cos \theta, \quad B_1 + B_3 = B_2,$$

$$(B_1 - B_3) \sqrt{K} \cos \phi = B_2 \sqrt{K'} \cos \theta \quad \text{and} \quad (A_1 + A_3) \sqrt{K} = A_2 \sqrt{K'}.$$

On solving for the four unknown quantities we obtain

$$A_3 = A_1 \frac{\sqrt{K'} \cos \phi - \sqrt{K} \cos \theta}{\sqrt{K'} \cos \phi + \sqrt{K} \cos \theta}, \quad B_3 = B_1 \frac{\sqrt{K} \cos \phi - \sqrt{K'} \cos \theta}{\sqrt{K} \cos \phi + \sqrt{K'} \cos \theta},$$

$$A_2 = A_1 \frac{2\sqrt{K} \cos \phi}{\sqrt{K} \cos \theta + \sqrt{K'} \cos \phi}, \quad B_2 = B_1 \frac{2\sqrt{K} \cos \phi}{\sqrt{K'} \cos \theta + \sqrt{K} \cos \phi}.$$

On substituting $\frac{\sin \phi}{\sin \theta}$ for $\sqrt{\frac{K'}{K}}$, these results become

$$A_3 = A_1 \frac{\tan (\phi - \theta)}{\tan (\phi + \theta)}, \quad B_3 = -B_1 \frac{\sin (\phi - \theta)}{\sin (\phi + \theta)},$$

$$A_2 = A_1 \frac{2 \sin \theta \cos \phi}{\sin (\phi + \theta) \cos (\phi - \theta)}, \quad B_2 = B_1 \frac{2 \sin \theta \cos \phi}{\sin (\phi + \theta)}.$$

The above are called Fresnel's formulæ. They were first obtained by Fresnel but by another method. They enable us to determine completely the reflected and refracted waves when the incident wave is known.

According to these formulæ B_3 never vanishes but A_3 becomes equal to zero when $\tan (\phi + \theta) = \infty$, i.e. when $\phi + \theta = \frac{\pi}{2}$. In this case $\sin \theta = \cos \phi$, and if n be put for the ratio of the refractive indices of both media, i e. if $n = \sqrt{K'/K}$,

$$n = \frac{\sin \phi}{\sin \theta} = \tan \phi.$$

This value of ϕ is called the polarising angle, and this equation states Brewster's law, of which an account has already been given in Chapter XI. After reflection at this angle of incidence, natural light is plane polarised in the plane of incidence.

Fresnel's formulæ can be verified very easily with a spectrometer fitted with two nicols with square ends, one attached to the collimator and the other attached to the telescope in front of its object glass. These nicols can be rotated respectively about the axes of the collimator and telescope, and are provided with divided circles for reading their positions. The collimator has a circular aperture instead of a slit. From Fresnel's formulæ,

$$\frac{B_3}{A_3} = -\frac{B_1}{A_1}\frac{\sin(\phi-\theta)\tan(\phi+\theta)}{\sin(\phi+\theta)\tan(\phi-\theta)} = -\frac{B_1}{A_1}\frac{\cos(\phi-\theta)}{\cos(\phi+\theta)}.$$

A_1/B_1 is the tangent of the angle which the plane of polarisation makes with the xz plane before reflection, and A_3/B_3 the tangent of the like angle after reflection. In the experiments of Jamin and Quincke A_1/B_1 was put equal to unity, that is, the polarising nicol was set with its principal plane at 45° to the xz plane, then A_3/B_3 was determined experimentally for different values of ϕ, and the results compared with those given by the formula. The agreement was very good, only in the neighbourhood of the polarising angle was there an appreciable difference between theory and experiment. This difference has been shown to be due to the boundary conditions not being accurate. In deriving the latter, we assumed that the value of the index of refraction changed discontinuously in passing from the one medium to the other. If we assume that the change takes place gradually within a region small in comparison with the wave-length of light, we obtain more elaborate boundary conditions, and from these can derive formulæ that represent the experimental results perfectly. From experiments confirming the more accurate theory, we learn that the transition layer or region in which the index of refraction changes from the one value to the other has, in the case of a polished glass surface, a thickness of about $\frac{2}{100}$ of the wave-length of sodium light.

Perpendicular Incidence. In the case of perpendicular incidence ϕ and θ both become zero and Fresnel's formulæ for A_3 and B_3 become indeterminate. If, however, we use the equations on p. 409 immediately above Fresnel's formulæ, $\cos\phi$ and $\cos\theta$ both become equal to 1, and

$$A_3 = A_1\frac{\sqrt{K'}-\sqrt{K}}{\sqrt{K'}+\sqrt{K}} = A_1\frac{n-1}{n+1}, \quad B_3 = B_1\frac{\sqrt{K}-\sqrt{K'}}{\sqrt{K}+\sqrt{K'}} = B_1\frac{1-n}{1+n}.$$

The fraction of the intensity reflected is therefore the same for light polarised in and perpendicular to the plane of incidence, namely $\left(\frac{n-1}{n+1}\right)^2$. In the case of reflection from glass to air, $n = 1{\cdot}5$ approximately; hence 4 per cent of the incident light is reflected.

Total Reflection. Suppose that K' is less than K, that the wave, for example, is reflected internally at a glass-air surface. Then ϕ is the angle of incidence in the glass, θ the angle of refraction in the air, and $\sin\theta = n\sin\phi$, n of course having its usual value of 1·5 or thereabouts. We have

$$\cos\theta = \sqrt{1 - \sin^2\theta} = \sqrt{1 - n^2\sin^2\phi}.$$

Where total reflection occurs, $n^2\sin^2\phi$ is greater than 1 and $\cos\theta$ becomes imaginary. We may write it in this case,

$$\cos\theta = i\sqrt{n^2\sin^2\phi - 1}.$$

It is interesting to examine what happens to Fresnel's formulæ when this imaginary value of $\cos\theta$ is substituted. Let us at first confine our attention to the reflected wave and examine the expression for B_3. For angles of incidence greater than the limiting angle,

$$B_3 = -B_1\frac{\sin(\phi - \theta)}{\sin(\phi + \theta)} = -B_1\frac{\sin\phi\cos\theta - \cos\phi\sin\theta}{\sin\phi\cos\theta + \cos\phi\sin\theta}$$

$$= -B_1\frac{i\sin\phi\sqrt{n^2\sin^2\phi - 1} - n\sin\phi\cos\phi}{i\sin\phi\sqrt{n^2\sin^2\phi - 1} + n\sin\phi\cos\phi}$$

$$= B_1\frac{n\cos\phi - i\sqrt{n^2\sin^2\phi - 1}}{n\cos\phi + i\sqrt{n^2\sin^2\phi - 1}}.$$

On multiplying both numerator and denominator by

$$n\cos\phi - i\sqrt{n^2\sin^2\phi - 1}$$

this gives

$$B_3 = B_1\frac{(n^2\cos^2\phi - n^2\sin^2\phi + 1) - 2in\cos\phi\sqrt{n^2\sin^2\phi - 1}}{n^2 - 1}.$$

The coefficient of B_1 is a complex quantity, the modulus of which is found by calculation to be 1 and the amplitude of which is

$$\tan^{-1}\frac{2n\cos\phi\sqrt{n^2\sin^2\phi - 1}}{n^2\cos^2\phi - n^2\sin^2\phi + 1}.$$

On writing b for the latter, the equation becomes

$$B_3 = B_1\, e^{-ib}.$$

In order to interpret this result it is necessary to go back somewhat.

$$Y_1 = B_1\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{K}(x\sin\phi + z\cos\phi)}{c}\right)$$

represented the electric intensity perpendicular to the plane of incidence for the incident wave. Instead of the cosine we might have written

$$Y_1 = \text{real part of } B_1 e^{i\frac{2\pi}{\tau}\left(t - \frac{\sqrt{K}(x\sin\phi + z\cos\phi)}{c}\right)}$$

and we could have made similar substitutions for the other cosines. This assumption is perfectly legitimate, for the equations of the electromagnetic field and the boundary conditions are linear in X, Y, Z, α, β, γ; they are satisfied by both parts of the complex quantities taken singly, and therefore must be satisfied by their sum. Had we proceeded in this way, we should have found for Y_3 in the above case,

$$Y_3 = \text{real part of } B_3 e^{i\frac{2\pi}{\tau}\left(t - \frac{\sqrt{K}(x \sin\phi - z\cos\phi)}{c}\right)}$$

$$= \text{real part of } B_1 e^{i\left[\frac{2\pi}{\tau}\left(t - \frac{\sqrt{K}(x \sin\phi - z\cos\phi)}{c}\right) - b\right]}$$

$$= B_1 \cos\left[\frac{2\pi}{\tau}\left(t - \frac{\sqrt{K}(x \sin\phi - z\cos\phi)}{c}\right) - b\right].$$

The amplitude of the reflected wave is therefore the same as the amplitude of the incident wave, i.e. no light is lost by reflection, but a phase difference is produced $= b$ and varying with the angle of incidence.

Similar results are obtained on examining the expression for A_3. Let us denote the phase difference produced in this case by a. Both components of the incident wave were originally in the same phase, but a relative phase difference has now grown up between them equal to $a - b$. This method of interpreting the complex amplitude is due to Fresnel.

The relative phase difference has been determined experimentally by a spectrometer of the type shown in fig. 307, and the results have

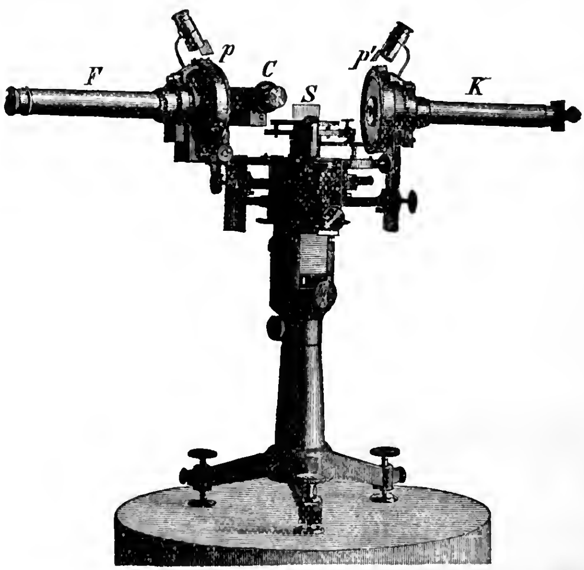


FIG. 307 (from Drude's "Optics").

K is the collimator, p' the polariser, **C** the compensator, p the analyser, **F** the telescope, and **S** the total reflection prism. The instrument depicted is the actual one used by Drude in determining the optical constants of metals.

been found to agree with theory. The collimator and telescope are fitted respectively with rotating polarising and analysing nicols the positions of which can be observed by reading microscopes. The collimator has a circular aperture instead of a slit. In front of the analyser is attached a Babinet compensator of the Soleil type. This compensator consists of two quartz wedges (fig. 308), in which the axis is perpendicular to the plane of the paper, and a quartz plate in which the axis is parallel to the plane of the paper in the direction shown; the quartz plate is cemented to one of the wedges. The whole arrangement forms a plate of variable thickness, which can produce a relative phase difference between the components polarised in and at right angles to the plane of the paper, and by turning the micrometer screw this phase difference can be varied from $-\lambda$ to $+\lambda$. The micrometer screw head can be seen in fig. 307. The difference between this compensator and the other type described on p. 224 is, that this one produces the same phase difference throughout the whole field, and hence is adapted for use with parallel light, whereas in the other type the phase difference alters as we move along the wedges.

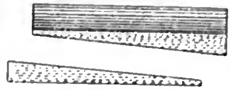


FIG. 308.

In verifying the theory the polarising nicol was set so as to polarise the light in a plane making an angle of 45° with the vertical. The light was then allowed to enter a total reflecting prism, be totally reflected by the hypotenuse face, and next to emerge from the third face and enter the compensator. The latter annulled the relative phase difference and the light was then extinguished by the analyser. From the angle between the directions of the telescope and collimator the angle of incidence of the light on the hypotenuse surface could be calculated.

When the compensator is removed, this apparatus can be used for verifying Fresnel's laws as described on p. 410.

Propagation of the Disturbance into the Second Medium. Let us now investigate what happens to the expression for the amplitude of the refracted wave when the angle of incidence exceeds the critical angle of total reflection. If we substitute

$$\sin\theta = n\sin\phi, \quad \cos\theta = i\sqrt{n^2\sin^2\phi - 1},$$

in A_2 and B_2, the expressions for the amplitudes, the result is not zero; if we substitute these values in

$$\cos\frac{2\pi}{\tau}\left(t - \frac{(x\sin\theta + z\cos\theta)}{c}\right),$$

the expression for the wave itself, we obtain

$$\cos\frac{2\pi}{\tau}\left(t - \frac{xn\sin\phi}{c} - i\frac{z}{c}\sqrt{n^2\sin^2\phi - 1}\right).$$

If we interpret this result in the same way as in the previous pages by writing an exponential in place of the cosine and then taking the real part, it becomes

$$e^{-\frac{2\pi z}{\lambda}\sqrt{n^2 \sin^2\phi - 1}} \cos\frac{2\pi}{\tau}\left(t - \frac{xn\sin\phi}{c}\right) \quad . \quad . \quad (20)$$

In order to make this result intelligible, we have taken the negative value of the root in the exponential. The exponential then diminishes very rapidly as z increases, and becomes inappreciable when z is 2 or 3 wave-lengths. The expression consequently represents a wave disturbance in the less dense medium close up to the surface and moving parallel to the x axis.

This result, namely, that when light is totally reflected there is nevertheless a wave in the second medium, appears at first astonishing but is borne out by experience. For if a convex surface of glass of large radius of curvature is brought into contact with the surface at which total reflection takes place, there is a transparent patch where the two surfaces are in optical contact, and this patch is surrounded by an edge, which transmits light of a reddish tint and reflects light of a bluish tint. At this edge the glass surfaces are not in contact but the air film is not thick enough to cause total reflection. Obviously according to (20) the red must penetrate furthest into the air; hence the tints of the transmitted and reflected lights. Although there is a wave in the second medium it is found on applying Poynting's theorem that there is no energy flowing into that medium; as long as the intensity of the incident wave is constant, all its energy reappears in the reflected wave. Another peculiarity of the wave in the second medium is that it is not transverse; on calculating out the expression for X it is found that it is not zero, and, of course, the wave is propagated in the x direction.

The question of the conditions in the second medium at total reflection has fascinated many experimenters from the time of Newton on. According to Newton's theory the light corpuscle passed through the glass in a straight line and was attracted by the glass when it entered the air, just in the same way as a stone projected into the air is attracted by gravity. The corpuscle consequently described a parabola in the air, then entered the glass again. According to Newton's reasoning, if another glass surface were gradually approached to the first, as soon as it touched the vertex of this parabola, some light should enter this second piece of glass, i.e. while there was still a thin air film between the two surfaces. Newton succeeded in getting this effect experimentally, though now we interpret it otherwise.

Propagation of Light in Absorbing Media. So far, in dealing with electromagnetic waves, we have confined ourselves to dielectrics. Let us now drop this restriction and assume that σ is not zero.

Then the equations of the field are

$$\frac{4\pi\sigma}{c}\mathsf{X} + \frac{\mathsf{K}}{c}\frac{\partial \mathsf{X}}{\partial t} = \frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}, \quad \frac{4\pi\sigma}{c}\mathsf{Y} + \frac{\mathsf{K}}{c}\frac{\partial \mathsf{Y}}{\partial t} = \frac{\partial\alpha}{\partial z} - \frac{\partial\gamma}{\partial x}, \quad \frac{4\pi\sigma}{c}\mathsf{Z} + \frac{\mathsf{K}}{c}\frac{\partial \mathsf{Z}}{\partial t} = \frac{\partial\beta}{\partial x} - \frac{\partial\alpha}{\partial y},$$

$$-\frac{\mu}{c}\frac{\partial\alpha}{\partial t} = \frac{\partial \mathsf{Z}}{\partial y} - \frac{\partial \mathsf{Y}}{\partial z}, \quad -\frac{\mu}{c}\frac{\partial\beta}{\partial t} = \frac{\partial \mathsf{X}}{\partial z} - \frac{\partial \mathsf{Z}}{\partial x}, \quad -\frac{\mu}{c}\frac{\partial\gamma}{\partial t} = \frac{\partial \mathsf{Y}}{\partial x} - \frac{\partial \mathsf{X}}{\partial y}.$$

Let us assume that we are dealing with harmonic plane waves of period τ, and that exponentials are to be substituted in place of X, Y, Z, α, β, γ, always on the understanding, of course, that the latter are the real parts of the exponential replacing them. Then, as t occurs in every quantity in the same factor $e^{i\frac{2\pi t}{\tau}}$, dividing by $i\frac{2\pi}{\tau}$ is equivalent to integrating with respect to t, and the first of the above equations may be written

$$\frac{1}{c}(K - i2\sigma\tau)\frac{\partial X}{\partial t} = \frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z}.$$

The second and third equations take the same form. If, as is usual in dealing with light waves, we put $\mu = 1$, the only effect of the conductivity of the medium is to replace K by the complex quantity $K - i2\sigma\tau$. The waves will therefore be represented by terms of the type

$$\text{real part } e^{i\frac{2\pi}{\tau}\left(t - \frac{\sqrt{K - i2\sigma\tau}(lx + my + nz)}{c}\right)}.$$

On writing $K - i2\sigma\tau = (\nu - i\kappa)^2$,

this becomes, real part $e^{i\frac{2\pi}{\tau}\left(t - \frac{(\nu - i\kappa)(lx + my + nz)}{c}\right)}$

$$= \text{real part } e^{-\frac{2\pi\kappa}{\tau c}(lx + my + nz)}\, e^{i\frac{2\pi}{\tau}\left(t - \frac{\nu(lx + my + nz)}{c}\right)}$$

$$= e^{-\frac{2\pi\kappa}{\tau c}(lx + my + nz)} \cos\frac{2\pi}{\tau}\left(t - \frac{\nu(lx + my + nz)}{c}\right).$$

This represents a wave the amplitude of which diminishes as the wave advances, the energy of which is being absorbed as it progresses. The exponential factor diminishes as $lx + my + nz$ increases. In a conductor we must have therefore absorption of electromagnetic waves. The planes of equal amplitude coincide with the planes of equal phase. The constant ν is termed the index of refraction of the absorbing medium and κ its absorption coefficient. κ usually varies rapidly with the wave-length.

Reflection at a Metal Mirror. Let a plane wave of light fall on a polished plane metal surface from air, and let the incident and reflected waves be represented by the same notation as on p. 408, except that K is here put $=1$, since the first medium is air. The amplitude of the refracted wave inside the metal dies away very rapidly. Let ν and κ be respectively the index of refraction and coefficient of absorption of the metal. It is seen from the preceding section that the propagation of a plane wave in an absorbing medium is formally the same as in a transparent medium, except that the complex expression $(\nu - i\kappa)^2$ takes the place of the specific inductive capacity. If we substitute this expression for K′ and put $K = 1$ all the results of p. 409 stand, only care must be taken in interpreting them.

Consider the cosine which occurs in all the expressions for the different components of the refracted wave, namely

$$\cos\frac{2\pi}{\tau}\left(t - \frac{\sqrt{K'}(x\sin\theta + z\cos\theta)}{c}\right).$$

It becomes

$$\cos\frac{2\pi}{\tau}\left(t - \frac{(\nu - i\kappa)(x\sin\theta + z\cos\theta)}{c}\right). \quad . \quad . \quad (21)$$

where $\sin\theta$ is given by

$$(\nu - i\kappa)\sin\theta = \sin\phi.$$

θ is consequently imaginary. We have

$$(\nu - i\kappa)^2(1 - \cos^2\theta) = \sin^2\phi,$$

$$(\nu - i\kappa)\cos\theta = \sqrt{(\nu - i\kappa)^2 - \sin^2\phi}.$$

Write $a - ib$ for the root on the right side of the equation ; a and b are of course functions of ϕ. Then

$$(\nu - i\kappa)\cos\theta = a - ib.$$

Substitute for $\cos\theta$ and $\sin\theta$ in (21) and it becomes

$$\cos\frac{2\pi}{\tau}\left(t - \frac{x\sin\phi + z(a - ib)}{c}\right)$$

$$= e^{-\frac{2\pi bz}{\tau c}}\cos\frac{2\pi}{\tau}\left(t - \frac{x\sin\phi + az}{c}\right).$$

The direction cosines of the normal to the surfaces of equal phase are obviously given by

$$\frac{\sin\phi}{\sqrt{a^2 + \sin^2\phi}},\ 0,\ \frac{a}{\sqrt{a^2 + \sin^2\phi}},$$

and the velocity of the refracted wave is

$$\frac{c}{\sqrt{a^2 + \sin^2\phi}}.$$

It varies with the angle of incidence.

The surfaces of equal amplitude are given by $z =$ const. and consequently do not coincide with the surfaces of equal phase except for the case of perpendicular incidence. In this case $\sin\phi = 0$, $a = \nu$, and $b = \kappa$. The amplitude, of course, gets rapidly less as z increases.

If we substitute the complex values of $\sin\theta$ and $\cos\theta$ in the expressions for A_3 and B_3 they also become complex. Both the components polarised in and at right angles to the plane of incidence thus suffer a change of phase on reflection, and the reflected wave is in general elliptically polarised, even if the incident wave is plane polarised. For two particular angles of incidence, however, perpendicular incidence and grazing incidence, there is no relative phase difference produced by reflection, and consequently, if the incident wave is plane polarised, the reflected wave is also plane polarised. The angle for which the relative phase difference is $\frac{1}{2}\pi$ is called the principal angle of incidence.

The optical constants of a metal can be determined with the polarisation spectrometer shown on p. 412 by observations on the positions of the analyser and compensator necessary to extinguish the reflected light, when the incident light is plane polarised at 45° to the vertical. Determinations were made by Drude according to this method, and it was found necessary to have the surface very carefully polished, as dirt, oxidation, etc., had a considerable influence on the results.

κ can be measured by having two thin metal films of different thickness and comparing the difference in the intensities of the transmitted beams. The reflection losses are then the same in both cases.

The index of refraction ν of metal prisms has been determined directly by Kundt. The prisms employed had an extremely small angle of refraction, less than one minute, and were prepared by the kathode discharge or by electrolytic deposition on platinised glass. When the incidence of the light on the prism is approximately perpendicular, the velocity of the refracted wave differs inappreciably from c/ν, and the deviation is given by the same formula as in the case of transparent media. Kundt's measurements give the striking result that for some metals ν is less than 1. The wave is consequently deviated towards the thin end of the prism. This is in agreement with Drude's determination by the other method.

When a plane polarised wave of light is incident from air perpendicularly on the plane surface of a transparent medium, the ratio of the amplitude of the reflected beam to the amplitude of the incident beam is given by

$$\frac{1-n}{1+n} \quad \text{or} \quad \frac{n-1}{n+1},$$

according as the incident beam is polarised in or at right angles to the plane of incidence. This result holds also for metallic mirrors provided that we write for n its complex equivalent $\nu - i\kappa$. The first of the expressions then becomes

$$\frac{1-\nu+i\kappa}{1+\nu-i\kappa}.$$

There is consequently a change of phase as well as a change of amplitude. If we multiply the expression by its conjugate

$$\frac{1-\nu-i\kappa}{1+\nu+i\kappa},$$

the result

$$\frac{(1-\nu)^2+\kappa^2}{(1+\nu)^2+\kappa^2} = \mathsf{R} = \frac{\nu^2+\kappa^2+1-2\nu}{\nu^2+\kappa^2+1+2\nu},$$

gives the square of the ratio of the amplitudes, or, in other words, the ratio of the intensities. This quantity has the same value for the other component; consequently the ratio of the intensities is always the same, no matter what the state of polarisation of the incident light is. In all metals 2ν is small compared with $1+\kappa^2$, hence R is very nearly 1, i.e. all metals possess a high power of reflection, which is higher the greater the coefficient of absorption κ. It is this high power of reflection and nothing more that we denote by the term "metallic lustre". Air bubbles below the surface of water have metallic lustre, when they are visible only by light which has been totally reflected at their surface.

Gold and copper appear yellow because κ is great for that colour and consequently light of that colour is strongly reflected. The surface colours of metals are approximately complementary to the colours they show by transmitted light. Thus gold leaf appears green by transmitted light. For metals κ varies from 2 to 4. For anilin colouring matters κ is also large enough to influence the reflection, and they also show "surface colour," but for the great majority of substances, copper sulphate and potassium bichromate crystals for example, κ is too small to influence the colour of the reflected light appreciably. These salts have what is known as body colour, i.e. the light penetrates some distance in before it is reflected, and during its passage in and out some of the constituents of white light are absorbed. The emergent light consequently appears coloured. In this case the colour of the reflected light is approximately the same as the colour of the transmitted light.

The fraction of light transmitted perpendicularly through a layer d cms. thick is

$$e^{-\frac{4\pi\kappa d}{\lambda}},$$

where λ is the wave-length in air. In dealing with solutions on p. 328 we represented the same quantity by

$$10^{-\mathrm{A}cd},$$

where A and c are constants. These two expressions are merely different ways of expressing the same thing. For example, the second may be written

$$e^{-2{\cdot}302\,\mathrm{A}cd}.$$

The base e is usual in theoretical work, while the base 10 is more convenient for practical determinations.

It should be mentioned that the quantities coefficient of absorption and molecular extinction coefficient are not universally recognised as defined here. The specification of the absorption of light is a matter on which there is still much confusion, and before using any results obtained in this field it is always necessary to find out in what units they are expressed.

Fresnel's Theory of Reflection. A theory of the reflection and refraction of light at the plane surface of a transparent medium, the results of which were in accordance with facts, was first given by Fresnel. This theory has the merits of straightforwardness and simplicity but did not meet with the approval of succeeding mathematicians, and theory succeeded theory throughout the greater part of last century until finally Hertz's experiments established the electromagnetic theory on a firm basis.

Fresnel regarded light waves as analogous to elastic waves. Now when elastic waves are propagated, e.g. sound waves in a gas or torsional vibrations up a stretched wire, the velocity of the wave is equal to the square root of the quotient of an elasticity modulus by a density. Fresnel assumed that all space was filled by a very light medium called the ether, which extended to the sun and stars and which penetrated inside all bodies, filling the spaces between their molecules. The light waves were elastic waves in this ether. The density of the ether was constant inside a medium, if the index of refraction was constant, but varied from medium to medium. The elasticity of the ether was constant throughout all media.

Suppose now that a plane wave is incident at an angle ϕ on the surface of a transparent medium, as shown in fig. 309, and gives rise to a reflected wave and a refracted wave, the angle of refraction being θ. For the sake of clearness we can suppose that the upper medium is air and the lower medium glass. Let v be the velocity of light in air and v' in glass, let ρ be the density of the ether in air and ρ' in glass, and let n be the index of refraction of glass. Then

$$\frac{v}{v'} = \sqrt{\frac{\rho'}{\rho}},$$

since the elasticity is the same for both media. But

$$\frac{v}{v'} = n.$$

Hence $$n = \sqrt{\frac{\rho'}{\rho}},$$

i.e. the index of refraction of glass is equal to the root of the quotient of the density of the ether inside glass by the density of the ether inside air.

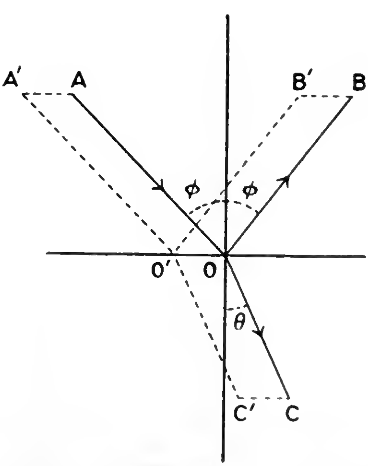


FIG. 309.

Let a be the amplitude of the incident wave, b the amplitude of the reflected wave, and c the amplitude of the refracted wave. The energy of the incident wave must be equal to the sum of the energies of the reflected and refracted waves. To see what relation this involves between a, b, and c, let OA (fig. 309) equal v, let OB $= v$, let OC $= v'$, let O′A′, O′B′, and O′C′ be respectively parallel and equal to OA, OB, and OC, and consider the energy in the parallelipiped standing on AO′ and of unit thickness in a direction normal to the paper. After unit time has elapsed the energy which was initially in AO′ has moved into BO′ and CO′. The energy per unit volume is proportional to the density of the ether and to the square of the amplitude. The volume occupied by the portion of the incident wave under consideration is AO × OO′ sin O′OA = vOO′ cos ϕ. This is also equal to the volume occupied by the reflected wave, while the refracted wave occupies v'OO′ cos θ. The equation of energy consequently gives

$$a^2\rho v OO' \cos\phi = b^2\rho v OO' \cos\phi + c^2\rho' v' OO' \cos\theta$$

or

$$(a^2 - b^2)\rho v \cos\phi = c^2\rho' v' \cos\theta.$$

Now $v/v' = n = \sin\phi/\sin\theta$ and $\rho/\rho' = 1/n^2 = \sin^2\theta/\sin^2\phi$. Substitute these values and the energy equation becomes

$$(a^2 - b^2)\sin\theta\cos\phi = c^2\sin\phi\cos\theta$$

or

$$\frac{a^2 - b^2}{c^2} = \frac{\tan\phi}{\tan\theta} \quad . \quad . \quad . \quad . \quad (22)$$

The ether in the two media can be regarded as two portions of different elastic substances in contact. They must remain in contact during the motion, i.e. there must be no slipping at the interface. The displacement at the interface must therefore be the same in both media.

Fresnel assumed that the vibration is at right angles to the plane in which the light is polarised. Consequently if the light is polarised in the plane of incidence, the complete displacement in the air is $a + b$ and in the glass is c. Thus

$$a + b = c \quad . \quad . \quad . \quad . \quad . \quad (23)$$

Divide (22) by (23) and we obtain

$$a - b = c\tan\phi\cot\theta.$$

Substitute for c from (23) and this gives

$$a - b = (a + b)\tan\phi\cot\theta,$$

which simplifies to

$$b = -a\frac{\sin(\phi - \theta)}{\sin(\phi + \theta)}.$$

Substitute this value for b in (23) and we obtain

$$c = a\frac{2\cos\phi\sin\theta}{\sin(\phi + \theta)}.$$

These expressions are the same as have already been obtained by the electromagnetic theory.

If the light is polarised at right angles to the plane of incidence,

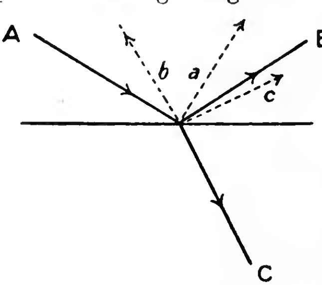


FIG. 310.

the vibrations make an angle with the trace of the surface, as is shown in fig. 310, and we take their components parallel to the surface.

The condition for equal displacement is then

$$(a - b) \cos \phi = c \cos \theta \qquad . \quad . \quad . \quad (24)$$

Divide (22) by this equation and we obtain

$$a + b = c \frac{\sin \phi}{\sin \theta}.$$

Substitute for c from (24) and we obtain

$$a + b = (a - b) \frac{\sin \phi \cos \phi}{\sin \theta \cos \theta}$$

which simplifies to $$b = a \frac{\tan (\phi - \theta)}{\tan (\phi + \theta)},$$

and on substituting this value for b in (24) we obtain

$$c = a \frac{2 \cos \phi \sin \theta}{\sin (\phi + \theta) \cos (\phi - \theta)},$$

results which we have already obtained by the electromagnetic theory.

The criticism directed against Fresnel's theory as given above was, that he did not employ the correct boundary conditions for the passage of an elastic wave from one medium to another. Not only should the tangential component of the displacement be the same on both sides of the boundary but the normal component should also be the same. This condition is clearly violated above in the case of the component polarised at right angles to the plane of incidence. To restore it we must assume that the refracted wave is not transverse but has a longitudinal component, i.e. the displacement has a component parallel to the direction of propagation. This longitudinal wave has theoretically a different velocity from the velocity of the transverse wave. All experimental evidence is against its existence, and it has proved an insurmountable obstacle to the rigorous elastic solid theory of light.

Wiener's Experiment. It has been explained on p. 151 that when light waves are incident perpendicularly on a reflecting surface, stationary waves are formed. The existence of these stationary waves was demonstrated directly in a very elegant manner by O. Wiener in 1890. He used a very thin photographic film, only $\frac{1}{30}$ wave-length of light thick, which was coated on a glass plate, and placed it almost parallel to the mirror, film side in, and with one edge touching the surface of the mirror. A silver chloride film of this thickness was perfectly transparent; consequently the stationary waves were not affected by its presence. Where a node intersected the film there was no photographic action, and where a loop intersected the film there was a maximum of photographic action. When the film was developed it was found blackened along the loops with the intervening portions clear. When the angle between the surface of the mirror and the film was increased, a greater number of loops was intersected in the same distance, and the black lines were found to be finer and closer together.

According to the electromagnetic theory, when light is reflected at perpendicular incidence from a metal mirror there should be nodes of the electric intensity approximately on (really slightly behind) the surface of the mirror and at distances of multiples of $\lambda/2$ out from this, while the nodes of the magnetic intensity should be approximately midway between these planes. Wiener found that the maxima of photographic action corresponded to the positions of the loops of electric intensity. The experiment was repeated by Drude and Nernst, who used a thin fluorescent film in place of the photographic film, and the loops of the electric intensity showed their presence by equidistant green lines. Thus it is the electric intensity in the light wave which causes both the photographic and fluorescent actions.

The same conclusion is supported by a further experiment due to Wiener in which light was incident on a metal mirror at an angle of 45° with the normal. When the light was plane polarised in the plane of incidence, the presence of stationary waves could be detected in front of the mirror by the method of the thin photographic film, but, when the incident light was plane polarised at right angles to the plane of incidence, there was no interference. In the first case the electric intensities of both the incident and reflected waves are at right angles to the plane of incidence and hence are in a condition to interfere, while in the second case they make a right angle with one another and interference is not possible.

Considerable importance used to be attached to the question as to whether the light vibrations took place in or at right angles to the plane in which they were polarised. According to the electromagnetic theory something takes place in both directions, the electric intensity is at right angles to and the magnetic intensity is in the plane in which the wave is polarised. This was stated as a fact in Chapter XXII, and we have seen from the results obtained for the polarising angle in this chapter, that the statement is in accordance with the definition introduced in Chapter XI. Wiener's experiments show, that it is the electric intensity which is responsible for the photochemical actions and fluorescence. As the action of light on the retina is probably photochemical in its nature, it too will be associated with the electric intensity.

EXAMPLES.

(1) Work out directly for perpendicular incidence the case of reflection at the plane surface of a transparent medium.

(2) The values of ν and κ, as determined experimentally by the polarisation spectrometer for sodium light, are given for seven different metals in the tables at the end of the book. Calculate the percentage of light reflected at normal incidence by these same metals, and compare the results in the cases of silver and platinum with the values determined directly by Hagen and Rubens.

(3) On p. 227 Fresnel's rhomb was described as a parallelipiped of crown glass, in which a beam of light is twice reflected internally at an angle of 55° after having entered normally through the end. If the rhomb is to be made of the hard crown glass, for which the constants are given on p. 61, calculate the value of the angles necessary to make the phase difference exactly right for the D lines, and find at the same time what the phase difference is for the C, F, and G lines.

(4) Graph A_3/A_1 and B_3/B_1 for crown glass as functions of the angle of incidence from 0° to 90°.

(5) Calculate the percentage of yellow light transmitted perpendicularly through a film of silver one-twentieth wave-length thick.

CHAPTER XXIV.

THE THEORY OF DISPERSION.

According to the electromagnetic theory as formulated by Maxwell the specific inductive capacity was a constant for the medium and the medium was regarded as continuous. As the index of refraction is equal to the square root of the specific inductive capacity, it follows that the index of refraction is also a constant for the medium, and consequently the electromagnetic theory in its original form is unable to explain the dispersion of light. This difficulty has been overcome by assuming that the structure of matter is not continuous, but that matter contains electrons and ions which vibrate under the action of light waves, and the explanation has been so successful that dispersion is now regarded as one of the strong points of the electromagnetic theory. We shall now proceed to investigate the effect on the velocity of light of the electrons in a medium.

Maxwell's first set of equations for the electromagnetic field was of the form

$$\frac{4\pi\sigma}{c}\mathsf{X} + \frac{\mathsf{K}}{c}\frac{\partial \mathsf{X}}{\partial t} = \frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z} \quad . \quad . \quad . \quad (25)$$

Let us suppose that we are dealing with a medium in which there is a great number of particles of the same mass m, each carrying the same charge $-e$, and that there are N of these particles per unit volume at the point under consideration. We shall leave over the question as to whether these particles are electrons or ions, i.e. charged atoms. Let $\frac{dx}{dt}$, $\frac{dy}{dt}$, and $\frac{dz}{dt}$ be the average velocities of the particles in the x, y, and z directions. The number of particles crossing unit area at right angles to the x direction per second is $\mathsf{N}dx/dt$, and they carry with them a charge $-\mathsf{N}e\,dx/dt$. But by the definition of conductivity this is equal to $\sigma\mathsf{X}$. Let us suppose that the particles are moving about in a medium of specific inductive capacity unity. Then on substituting for $\sigma\mathsf{X}$ and putting $\mathsf{K} = 1$ equation (25) becomes

$$-\frac{4\pi\mathsf{N}e}{c}\frac{dx}{dt} + \frac{1}{c}\frac{\partial \mathsf{X}}{\partial t} = \frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z} \quad . \quad . \quad . \quad (26)$$

Let the charged particles oscillate about equilibrium positions. Then each is subject to an equation of the form

$$m\frac{d^2x}{dt^2} + h\frac{dx}{dt} + fx = 0 \quad . \quad . \quad . \quad . \quad (27)$$

where x denotes the displacement of the particle from its equilibrium position. The first term gives the rate of increase of momentum of the particle, and the third term gives a force attracting it towards its equilibrium position and proportional to its distance from that position; the second term gives a force proportional to the velocity, which always resists the motion. The force represented by the third term is analogous to the component of the weight of the bob of a pendulum, and the force represented by the second term to the resistance due to air friction.

To solve this equation substitute

$$e^{pt}$$

for x. This gives

$$mp^2 + hp + f = 0,$$

the roots of which are

$$p = \frac{-h \pm \sqrt{h^2 - 4mf}}{2m}.$$

Let us assume that the quantity inside the root is negative. Then the solution of (27) is

$$x = \mathsf{A}e^{-\frac{ht}{2m}} \cos\left\{\frac{\sqrt{4mf - h^2}}{2m}t - \alpha\right\} \quad . \quad . \quad . \quad (28)$$

where A and α are constants. The solution represents a damped simple harmonic motion, the period of which is given by

$$\tau = \frac{4\pi m}{\sqrt{4mf - h^2}}.$$

This period is referred to as the "free period"; h^2 is regarded as small in comparison with $4mf$ so that the free period is approximately

$$2\pi\sqrt{\frac{m}{f}} \quad . \quad . \quad . \quad . \quad . \quad (29)$$

When the light wave is passing, each particle is acted on by a force $-e\mathsf{X}$ due to the electric intensity of the wave. Instead of equation (27) we have therefore

$$m\frac{d^2x}{dt^2} + h\frac{dx}{dt} + fx = -e\mathsf{X} \quad . \quad . \quad . \quad (30)$$

Let us suppose that the light wave is a harmonic one. The equation then represents forced vibrations. The solution consists of two parts, one exactly the same as (28), called the free vibration, and another having the same period as the incident wave. This second part is called the forced vibration. The free vibration dies down rapidly owing to the exponential, so we have to consider only the forced vibration.

Since equation (30) is linear, we can write exponentials instead of sines or cosines. Assume that the period of the incident wave is $2\pi/g$. Then t occurs in x solely in the factor e^{igt} and a differentiation of x with regard to t is equivalent to a multiplication by ig. Thus equation (30) reduces to

$$(-mg^2 + ihg + f)x = -e\mathsf{X}.$$

So far x has been the coordinate of a representative particle. By writing down the similar equations for all the particles in the unit volume we find, that if x denotes the average displacement of the particles,

$$(-mg^2 + ihg + f)\mathrm{N}ex = -\mathrm{N}e^2\mathrm{X}.$$

Differentiate with respect to t and bring the bracket on the left to the denominator on the right. Then

$$\mathrm{N}e\frac{dx}{dt} = \frac{-\mathrm{N}e^2}{(-mg^2 + ihg + f)}\frac{\partial \mathrm{X}}{\partial t}.$$

Now substitute in (26) and the latter equation becomes

$$\frac{1}{c}\left(1 + \frac{4\pi\mathrm{N}e^2}{f - mg^2 + ihg}\right)\frac{\partial \mathrm{X}}{\partial t} = \frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}.$$

This is formally the same as the equation for the propagation of light in absorbing media. We have, therefore,

$$(\nu - i\kappa)^2 = 1 + \frac{4\pi\mathrm{N}e^2}{(f - mg^2) + ihg} \quad . \quad . \quad (31)$$

If a wave is propagated, then, in a medium which contains electrons vibrating in the above manner, its index of refraction and coefficient of absorption are given by the above equation. The y and z components of the vibrations of the electrons, when treated in the same way, lead to the same result.

The derivation of (31) is difficult to follow, but its application to the results of experiment is comparatively simple.

"Normal" Dispersion. Let us suppose that in the denominator of (31) hg is so small that it can be neglected in comparison with $(f - mg^2)$. Then the right-hand side of the equation becomes real, consequently κ becomes 0, and the medium is transparent. Equation (31) thus simplifies to

$$\nu^2 = 1 + \frac{4\pi\mathrm{N}e^2}{f - mg^2} \quad . \quad . \quad . \quad . \quad . \quad (32)$$

If λ is the wave-length of the light in vacuo

$$\lambda = \frac{2\pi c}{g}.$$

Write $m/f = (\lambda_0/2\pi c)^2$. We see from (29) that λ_0 is the wave-length corresponding to the free period of the electrons. Then (32) can be written

$$\begin{aligned}\nu^2 &= 1 + \frac{4\pi\mathrm{N}e^2}{m(2\pi c)^2\left(\frac{1}{\lambda_0^2} - \frac{1}{\lambda^2}\right)} \\ &= 1 + \frac{\mathrm{N}e^2\lambda_0^2\lambda^2}{m\pi c^2(\lambda^2 - \lambda_0^2)} \\ &= 1 + \frac{\mathrm{M}\lambda^2}{\lambda^2 - \lambda_0^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)\end{aligned}$$

where $M = Ne^2\lambda_0^2/(m\pi c^2)$. Let us assume that λ_0 lies in the ultra-violet and write $_v$ for the suffix $_0$. Then λ_v^2 is less than λ^2, and the denominator can be expanded in powers of λ_v^2/λ^2. Thus

$$\nu^2 = 1 + M\left(1 - \frac{\lambda_v^2}{\lambda^2}\right)^{-1}$$

$$= 1 + M + \frac{M\lambda_v^2}{\lambda^2} + \frac{M\lambda_v^4}{\lambda^4} \quad . \quad . \quad . \quad (34)$$

The dispersion of most transparent substances can be represented fairly well by the above formula, stopping at the λ^{-4} term. A better agreement can be obtained by assuming that, in addition to the electrons with the free period in the ultra-violet, there are also electrons with their free period in the infra-red. The effect of the latter is to add a new term to the expression for ν^2, which becomes

$$\nu^2 = 1 + \frac{M_v\lambda^2}{\lambda^2 - \lambda_v^2} + \frac{M_r\lambda^2}{\lambda^2 - \lambda_r^2}.$$

M_v refers to the electrons with the period in the ultra-violet, and M_r and λ_r to the electrons with the period in the infra-red. If we assume that λ_r is much greater than λ, the new term reduces simply to $-M_r\lambda^2/\lambda_r^2$, and the formula for the index of refraction becomes

$$\nu^2 = -\frac{M_r\lambda^2}{\lambda_r^2} + 1 + M_v + \frac{M_v\lambda_v^2}{\lambda^2} + \frac{M_v\lambda_v^4}{\lambda^4}$$

or

$$\nu^2 = -A'\lambda^2 + A + B\lambda^{-2} + C\lambda^{-4} \quad . \quad . \quad . \quad (35)$$

To illustrate the use of this formula I have calculated out the values of the constants for the case of the index of refraction of water. The first column in the following table gives wave-lengths, and the second the observed values of the index of refraction of water corresponding to these wave-lengths.

	$\lambda\ 10^5$.	ν.	ν^2 Observed.	ν^2 Calculated.	Difference 10^4.
A	7·60	1·3293	1·7670	1·7672	− 2
B	6·87	1·3309	1·7713	1·7713	0
C	6·563	1·3317	1·7734	1·7733	+ 1
D	5·893	1·3335	1·7782	1·7783	− 1
E	5·270	1·3358	1·7845	1·7843	+ 2
F	4·861	1·3377	1·7894	1·7894	0
f	4·341	1·3410	1·7982	1·7980	+ 2
G	4·308	1·3412	1·7988	1·7987	+ 1
H	3·968	1·3441	1·8066	1·8066	0

Four Fraunhofer lines A, D, F, and H, which are roughly equidistant from one another in the visible spectrum, were taken, and the wave-length and observed value of ν corresponding to each of these lines substituted in the formula. This gave four equations for the four unknown quantities A', A, B, and C; these equations when solved gave

$$A' = 1{\cdot}38\ 10^6,\ A = 1{\cdot}7642,\ B = 6{\cdot}12\ 10^{-11},\ \text{and}\ C = 1{\cdot}41\ 10^{-20}$$

These values were then substituted in the formula and ν^2 calculated. The results are given in the fourth column. The third column gives the observed value of ν^2, and the fifth column the difference of the observed and calculated values. The difference might possibly be made smaller by using additional places of decimals in the calculation.

As has already been mentioned on p. 401 the dielectric constant is the value which ν^2 takes when λ is infinite. When λ is very great, λ_r/λ becomes smaller than 1. The infra-red term

$$\frac{\mathsf{M}_r\lambda^2}{\lambda^2 - \lambda_r^2}$$

can then be expanded in the same way as the ultra-violet one and gives

$$\mathsf{M}_r\left(1 + \frac{\lambda_r^2}{\lambda^2} + \frac{\lambda_r^4}{\lambda^4}\ \cdot\cdot\cdot\right).$$

If now we make λ infinite, ν^2 becomes equal to K and all the terms in the expansions vanish except the constant terms. We have therefore

$$\mathsf{K} = 1 + \mathsf{M}_v + \mathsf{M}_r.$$

But A, the constant term in (35), is equal to $1 + \mathsf{M}_v$. Hence

$$\mathsf{K} - \mathsf{A} = \mathsf{M}_r.$$

If we take $\mathsf{K} = 80$ and substitute the value for A, this gives approximately

$$\mathsf{M}_r = 78.$$

But
$$\mathsf{A}' = \mathsf{M}_r/\lambda_r^2 = 1{\cdot}38\ 10^6.$$

Hence
$$\lambda_r^2 = \frac{\mathsf{M}_r}{\mathsf{A}'} = \frac{78}{1{\cdot}38\ 10^6},$$

which gives $\lambda_r = 7{\cdot}5\ 10^{-3}$ cms. It will be shown further down that there is an absorption band in the spectrum at every point where the electrons have a free period. Consequently, according to our calculation, there should be an absorption band at $\lambda = 7{\cdot}5\ 10^{-3}$ or 75μ. This is too far in the infra-red to be detected by a fluorite or rock-salt prism and bolometer. Water has however very many strong absorption bands in the near infra-red. The assumption made in the theory, namely, that there is only one free period in the infra-red, is too simple for a close agreement to be expected. But in any case the theory explains satisfactorily the difference between ν^2 and K.

If we consider the coefficients in the formula due to the ultra-violet free period, we find that theoretically $(\mathsf{A} - 1)\ \mathsf{C} = \mathsf{B}^2$, since each equals $\mathsf{M}_v^2\,\lambda_v^4$, but that this relation is not fulfilled. There is probably more than one free period in the ultra-violet. If, in order to obtain a rough result, we adhere to our assumption of one free period,

$$\lambda_v^2 = \frac{\mathsf{B}}{\mathsf{A} - 1} = \frac{6{\cdot}12\ 10^{-11}}{{\cdot}7642},$$

i.e.
$$\lambda_v = 8{\cdot}9\ 10^{-6}\ \text{cm}.$$

It has not been possible to investigate the absorption spectrum of water as far as this in the ultra-violet, but as far as water has been investigated it is transparent.

We have
$$M_v = \frac{Ne^2\lambda_v^2}{m\pi c^2} = \cdot 764.$$
Substitute $8 \cdot 9\ 10^{-6}$ for λ_v and $1 \cdot 772\ 10^7 c$ for e/m. $1 \cdot 772\ 10^7$ is the generally accepted value for the ratio of the charge to the mass of an electron in electromagnetic units, but in our notation e/m is measured in electrostatic units and hence the value has to be multiplied by c. We obtain then
$$\frac{Nm}{\pi}(1 \cdot 772\ 10^7 \times 8 \cdot 9\ 10^{-6})^2 = \cdot 764,$$
or
$$Nm = 9 \cdot 6\ 10^{-5}.$$
Nm is the total mass of the ultra-violet electrons per unit volume. m is $\frac{1}{1000}$ of the mass of a hydrogen atom, and hence $\frac{1}{18000}$ of the mass of a water molecule. If we suppose that there is one ultra-violet electron in each molecule, the total mass of the molecules per unit volume would be
$$9 \cdot 6\ 10^{-5} \times 18000 = 1 \cdot 7.$$
But the density of water is 1. Hence there are 1·7 electrons per molecule having a free period in the ultra-violet.

Interesting calculations of this nature have been made by Drude. He believed that the vibrating particles in the ultra-violet were electrons and those in the infra-red positive ions, and that the numbers of vibrating electrons and ions per molecule were each equal to the number of bonds in the molecule. Thus the 1·7 above would be an approximation for 2. His results are suggestive but by no means final. The theory of dispersion has taken us a certain distance, but it is impossible to get further until we have more knowledge of the structure of the molecule.

It should be stated, that for representing the indices of refraction of transparent substances in the visible spectrum Hartmann's empirical formula (p. 245) is better than the theoretical formula for ν^2. But in the neighbourhood of the absorption bands Hartmann's formula is of no use at all; also in the case of transparent substances, when the ultra-violet as well as the visible is considered, the theoretical formula is the better one.

§ Gladstone and Dale found experimentally that if μ is the refractive index of a gas and d its density, then
$$\frac{\mu - 1}{d}$$
is constant for the gas, no matter what the pressure is. In 1880 L. Lorenz in Copenhagen and H. A. Lorentz in Leyden showed on theoretical grounds that
$$\frac{\mu^2 - 1}{(\mu^2 + 2)d}$$
ought to be a constant for a substance. Experiment showed that this second expression enabled the refractive index of a vapour to be calculated when that of its liquid was known. Thus the expression takes the values 0·2068 and 0·2061 respectively for water vapour and water, 0·2898 and 0·2805 respectively for carbon disulphide vapour and carbon disulphide, and

0·1796 and 0·1790 for chloroform vapour and chloroform. The second expression can be written

$$\frac{\mu - 1}{d} \cdot \frac{\mu + 1}{\mu^2 + 2},$$

and, as the second factor in it does not vary rapidly with μ for small changes in μ, it must obviously lead to the same result as Gladstone and Dale's expression.

The physical chemists have applied the name specific refraction to both of these expressions indifferently, and have introduced the terms atomic and molecular refractions to denote the product of the specific refraction by the atomic weight and the molecular weight of the substance. They have found that the molecular refraction of a substance can be obtained roughly by adding the atomic refractions of the atoms composing its molecule. Or, in other words, molecular refraction is additive, but not always ; the constitution of the molecule has an influence on the accuracy of the agreement.

We have from equation (33)

$$\nu^2 - 1 = \frac{\mathsf{M}\lambda^2}{\lambda^2 - \lambda_0^2},$$

where M and λ_0 are constants. Write μ for ν and suppose that we have several different classes of electrons in the substance. Then

$$\mu^2 - 1 = \Sigma \frac{\mathsf{M}\lambda^2}{\lambda^2 - \lambda_0^2},$$

where each different class of electrons contributes its own term on the right-hand side. But for a given wave-length the right-hand side is constant. If the substance expands, all the quantities M vary inversely as d, for they all contain the number of electrons per cubic centimetre as a factor. Hence the theory of dispersion leads to the result known as Newton's law, namely, that it is

$$\frac{\mu^2 - 1}{d}$$

that should be constant, and neither of the previous expressions. Of course for small changes of μ Newton's law gives the same results as the other two formulæ.

"Anomalous" Dispersion. So far we have dealt only with transparent bodies. Let us now return to (31) and assume that κ is not zero. In order to simplify our results write, as before,

$$g = 2\pi c/\lambda, \; f/m = (2\pi c/\lambda_0)^2, \text{ and } \mathsf{M} = \mathsf{N}e^2\lambda_0^2/(m\pi c^2).$$

Then (31) becomes

$$(\nu - i\kappa)^2 = 1 + \frac{\mathsf{M}\lambda^2}{(\lambda^2 - \lambda_0^2) + i\mathsf{G}\lambda} \qquad (36)$$

where G is put for $\frac{h\lambda_0^2}{2m\pi c}$. On separating the real and imaginary parts (36) resolves into

$$\nu^2 - \kappa^2 = 1 + \frac{\mathsf{M}\lambda^2(\lambda^2 - \lambda_0^2)}{(\lambda^2 - \lambda_0^2)^2 + \mathsf{G}^2\lambda^2}$$

and

$$2\nu\kappa = \frac{\mathsf{MG}\lambda^3}{(\lambda^2 - \lambda_0^2)^2 + \mathsf{G}^2\lambda^2}.$$

Let us confine our attention now to solutions of inorganic salts, such as cobalt chloride or didymium chloride, solutions of anilin colouring matters, such as cyanine or fuchsine, and glasses coloured with, for example, cobalt oxide. In these cases we have one or more bands in the absorption spectrum, and the index of refraction varies in the neighbourhood of each band; κ^2, though, can always be neglected in comparison with ν^2, and ν does not vary much from the value which it would have for the solvent alone or for the glass without the addition of the colouring oxide.

In the expression for $2\nu\kappa$, as λ varies, the $(\lambda^2 - \lambda_0^2)^2$ term changes its value much more rapidly than the other terms. Hence we may write λ_0 for λ in the other terms and consider them constant. Divide both sides by 2ν. We have then

$$\kappa = \frac{MG\lambda_0^3}{2\nu((\lambda^2 - \lambda_0^2)^2 + G^2\lambda_0^2)} \quad . \quad . \quad . \quad (37)$$

To the same degree of approximation

$$\nu^2 = 1 + \frac{M\lambda_0^2(\lambda^2 - \lambda_0^2)}{(\lambda^2 - \lambda_0^2)^2 + G^2\lambda_0^2} \quad . \quad . \quad . \quad (38)$$

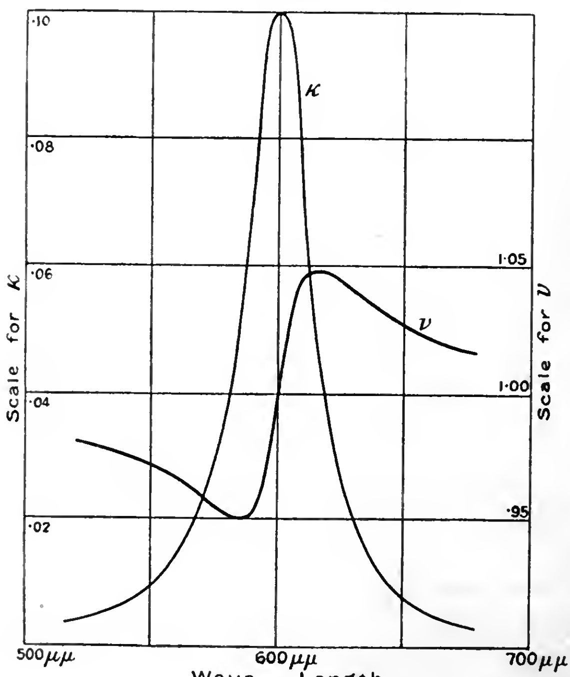


FIG. 311.

To illustrate the variation of ν and κ with λ they have both been plotted in fig. 311 as functions of λ for $M = \cdot 01$, $G = 3\ 10^{-6}$, and $\lambda_0 = 6\ 10^{-5}$

cms. It is seen that κ has a well-defined maximum at $\lambda = 6\ 10^{-5}$ cms., and that it falls rapidly to zero on both sides of this maximum, while ν decreases as we approach the band from the side of smaller wave-lengths, increases rapidly inside the band, and then decreases on the other side.

So far we have assumed only one free period. This is, of course, an ideal case. But in the case of well-defined bands in the visible spectrum with transparent regions between them the above formulæ can be applied, because, though each band has its own free period and although there are free periods in the ultra-violet causing the dispersion of the solvent, still throughout any one band the effect of the other bands on κ is negligible and their effect on ν is constant. Thus, to take a concrete case, if we are dealing with a solution of cyanine in alcohol, in the neighbourhood of the cyanine band

$$\nu^2 = 1 + \frac{M\lambda_0^2(\lambda^2 - \lambda_0^2)}{(\lambda^2 - \lambda_0^2)^2 + G^2\lambda_0^2} + \Sigma\frac{M'\lambda'^2_0(\lambda^2 - \lambda'^2_0)}{(\lambda^2 - \lambda'^2_0)^2 + G'^2\lambda'^2_0}$$

and

$$2\nu\kappa = \frac{MG\lambda_0^3}{(\lambda^2 - \lambda_0^2)^2 + G^2\lambda_0^2}.$$

The summation term gives the effect of all the ultra-violet and infra-red electrons, and, if we put n for the index of refraction of the solvent alone, the first of the above equations may be written

$$\nu^2 = n^2 + \frac{M\lambda_0^2(\lambda^2 - \lambda_0^2)}{(\lambda^2 - \lambda_0^2)^2 + G^2\lambda_0^2}.$$

According to the curve shown on p. 430 the index of refraction for $\lambda = 550\mu\mu$ is smaller than the index of refraction for $\lambda = 650\mu\mu$, i.e. the index of refraction is smaller on the violet side of the band. Consequently, if we fill a hollow prism with an alcoholic solution of cyanine, which has a well-marked absorption band at $590\mu\mu$, and use this prism to produce a spectrum, the red is more deviated than the violet and the colours come in the wrong order. This phenomenon is known as anomalous dispersion, though there is nothing anomalous about it. All substances which show normal dispersion in the visible spectrum must show anomalous dispersion in the ultra-violet. The increase in the index of refraction towards the violet in normal dispersion is merely the beginning of the right-hand slope of the ν curve in fig. 311 and is caused by the ultra-violet absorption. Hence instead of anomalous dispersion the name selective dispersion has been introduced to describe the phenomenon.

§ Selective dispersion was first properly studied by Christiansen and Kundt and is shown by all substances which have an absorption band in the visible spectrum. It is difficult to exhibit experimentally, because owing to the absorption band the prism must have a very small angle. Pflüger verified the theory by using prisms of solid cyanine of very small angle, determining ν with a spectrometer and measuring κ with a spectrophotometer. Thin parallel sided films were used with the spectrophotometer. The curves had not quite the shape demanded by theory for a single free period, but the agreement could be made perfect by assuming that the cyanine band consisted really of several bands close together superimposed.

Perhaps the simplest way of showing selective dispersion is as follows: An ordinary spectrometer is taken, a crown glass prism placed on the table, and the collimator slit replaced by a small circular diaphragm. Or instead of this the V-shaped piece that regulates the height of the slit is pushed in, until the height of the slit is no greater than its breadth. The slit is then illuminated with a metal filament lamp, and, as the source is practically a point one, a horizontal line spectrum is seen in the field of the telescope. The lamp is moved about until a portion of a filament comes on to the slit and the spectrum is as bright as possible. Next two square glass plates of about 2 or 3 cms. side are taken and a hollow prism of about 2° refracting angle made. This is easily done by placing the plates together with a wedge between them at the thick end of the prism, holding them in position with two elastic bands and making the sides tight with seccotine. After the seccotine is firm the elastic bands and wedge are removed. Care must be taken that no seccotine gets inside the prism at the refracting edge, as owing to the absorption the prism must be used where it is thinnest. Seccotine dissolves in acetic acid, and flows out of its tube better when slightly warmed.

When this prism is ready, it is filled with a very concentrated alcoholic solution of cyanine and placed on the spectrometer with its refracting edge horizontal and pointing downwards. It thus gives an upward deviation to the rays after they leave the crown glass prism. This deviation is, however, so small, that the spectrum is still in the field of the telescope usually about a quarter of a diameter of the field from the top. If the hollow prism were filled with alcohol, the spectrum would be a straight line. Owing, however, to the cyanine the yellow is absorbed and the index of refraction of the yellowish-red increased and of the yellowish-green diminished. Consequently the deviation of the yellowish-red is greater and of the yellowish-green less than it would be were there no cyanine in solution, and the line spectrum is broken into two parts as shown in fig. 312 which represents the appearance in the field. The positions of the different colours, red, orange, green, blue, indigo, violet, are indicated by the letters R, O, G, B, I, and V. The yellow is entirely absorbed. With the average spectrometer eyepiece under the conditions described above, if the average vertical deviation of the spectrum is $\frac{1}{4}$ diameter, the deviations of the orange and green just where they cease to be visible are respectively about $\frac{1}{40}$ diameter more and less than this. This arrangement for showing selective dispersion employs what is known as the method of crossed prisms.

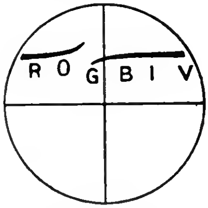


Fig. 312.

When κ^2 can be neglected in comparison with ν^2 as in the case of solutions and coloured glass, the absorption band does not appreciably affect the reflecting power of the substance. In the case of solid cyanine and fuchsine, however, κ^2 cannot be neglected in comparison with ν^2. Consequently (cf. p. 418) these substances reflect the colours which they absorb, or, in other words, they possess "surface colour".

It has been found by the method of residual rays (cf. p. 304) that quartz, fluorite, rock salt, and sylvin all possess well-marked regions of metallic reflection in the infra-red. Their indices of refraction in the infra-red can all be represented extremely well by the theoretical formula on the assumption that there are free periods at these regions.

Calculation of the Number of Electrons Causing an Absorption Band per Molecule. It was mentioned on p. 428 that Drude has made calculations from the constants of the dispersion formula as to the number of vibrating electrons there are in the ultra-violet per molecule for different substances. I have shown * that similar calculations can be made from the variation of κ throughout an absorption band in a solution of a dye. For in the formula (37) write n the index of refraction of the solvent for ν; then

$$\kappa = \frac{\mathsf{M}\mathsf{G}\lambda_0^3}{2n((\lambda^2 - \lambda_0^2)^2 + \mathsf{G}^2\lambda_0^2)}.$$

Let κ_m be the maximum value of κ; then

$$\kappa_m = \frac{\mathsf{M}\lambda_0}{2n\mathsf{G}} \quad . \quad . \quad . \quad . \quad . \quad (39)$$

Let λ_1 be the wave-length for which κ has half its maximum value; then λ_1 is given by

$$\lambda_1^2 - \lambda_0^2 = \pm \mathsf{G}\lambda_0$$

or

$$2(\lambda_1 - \lambda_0) = \mathsf{G}, \quad . \quad . \quad . \quad . \quad . \quad (40)$$

if we take the greater value of λ_1 and write $\lambda_1 + \lambda_0$ approximately equal to $2\lambda_0$. From (39) we have

$$\mathsf{M} = \frac{2n\kappa_m\mathsf{G}}{\lambda_0} = \frac{4n\kappa_m(\lambda_1 - \lambda_0)}{\lambda_0}$$

on substituting for G from (40). Now

$$\mathsf{M} = \frac{\mathsf{N}e^2\lambda_0^2}{m\pi c^2},$$

where m is the mass of the electron and N the number of electrons per unit volume. Suppose that e is measured in electromagnetic units; then the c^2 in the denominator disappears. Hence

$$\frac{\mathsf{N}e^2\lambda_0^2}{m\pi} = \frac{4n\kappa_m(\lambda_1 - \lambda_0)}{\lambda_0} \quad . \quad . \quad . \quad (41)$$

Let p be the number of electrons per molecule of colouring matter belonging to the absorption band under consideration. Then N/p is the number of molecules per unit of volume. Let s be the strength of the solution in gramme-molecules per litre, and let m_{H} be the mass of

* "Proc. Roy. Soc.," A **82**, p. 606, 1909.

an atom of hydrogen. Then $s/(1000m_{\mathsf{H}})$ is also equal to the number of molecules of dissolved colouring matter per unit volume. Therefore

$$\frac{\mathsf{N}}{p} = \frac{s}{1000m_{\mathsf{H}}}.$$

Substitute for N in (41) and we obtain

$$\frac{ps}{1000m_{\mathsf{H}}}\frac{e^2\lambda_0^{\,2}}{m\pi} = \frac{4n\kappa_m(\lambda_1 - \lambda_0)}{\lambda_0}$$

which gives

$$\frac{pe}{m} = 4000\pi\frac{n\kappa_m}{s}\frac{m_{\mathsf{H}}}{e}\frac{(\lambda_1 - \lambda_0)}{\lambda_0^{\,3}} \quad . \quad . \quad . \quad (42)$$

The following table shows some results obtained by the use of this formula. The first column gives the solution and the second and third the values of λ_0 and λ_1 used :—

Solution.	λ_0.	λ_1.	pe/m.
	$\mu\mu$	$\mu\mu$	
Fuchsine in alcohol	550	585	$1{\cdot}8\ 10^7$
,, ,, aniline	565	595	$1{\cdot}7\ 10^7$
Phloxin in water	515	560	$1{\cdot}4\ 10^7$
Crystal violet in alcohol . . .	575	611	$4{\cdot}9\ 10^7$
,, ,, ,, aniline . . .	605	635	$4{\cdot}7\ 10^7$
Corallin in alcohol	465	515	$1{\cdot}6\ 10^6$
,, ,, aniline	455	505	$1{\cdot}7\ 10^6$
Methylene blue in water . . .	665	690	$5{\cdot}4\ 10^6$
,, ,, ,, aniline . .	675	660	$7{\cdot}7\ 10^6$
Water blue in water	575	646	$8{\cdot}1\ 10^6$
Eosin in water	515	533	$6{\cdot}9\ 10^6$

As the generally accepted value of e/m is $1{\cdot}772\ 10^7$, the above table shows that probably one electron in each molecule is concerned in the production of the absorption band. When the formula is applied to solutions of inorganic salts such as cobalt chloride or didymium chloride the values of pe/m are very much less. In one case, potassium permanganate, the value is too great for an ion; hence the absorption bands of the inorganic salts are probably due to the vibrations of electrons, but only a very small proportion of the molecules possess such electrons.

Although the results given in the table are certainly significant, too much weight must not be placed on the actual numerical values obtained. The bands have never the simple shape demanded by theory but are always to some extent irregular. Also equation (30), on which the whole theory is based, namely,

$$m\frac{d^2x}{dt^2} + h\frac{dx}{dt} + fx = -e\mathsf{X},$$

probably does not satisfactorily represent the motion of an electron. The second term is open to question. When we get down to molecular dimensions, we cannot talk of frictional forces proportional to the velocity. When impacts between molecules occur, free vibrations are set up and energy is lost, but this loss of energy cannot be represented accurately in an equation and the second term is only an approximate way of allowing for it. Again the equation (30) holds only on the assumption that the electron is not acted on by the other electrons in its neighbourhood. If the assumption does not hold, more complicated systems will arise, of which it has as yet been impossible to form any definite picture.

The Dispersion of Metals. According to p. 415 if a harmonic wave of period τ is propagated through a medium of specific inductive capacity K and conductivity σ, the result is analytically the same as if the medium had a complex specific inductive capacity $K - i2\sigma\tau$. This specific inductive capacity was written $(\nu - i\kappa)^2$, where ν was the index of refraction and κ the coefficient of absorption of the medium. On equating the imaginary parts of these two expressions we obtain

$$\nu\kappa = \sigma\tau \quad . \quad . \quad . \quad . \quad . \quad (43)$$

Now, if the medium in question is a metal, σ is a constant and can be determined by electrical methods, and ν and κ can be determined by optical methods for light of any particular period τ. It is found, however, when we substitute the values in (43), that the equation is not satisfied, although it ought to be according to the original electromagnetic theory. But, just as in the case of transparent media, the discrepancy is removed when we drop the assumption that the structure of the medium is continuous.

Maxwell's first three equations of the electromagnetic field then take the form (cf. p. 423)

$$-\frac{4\pi Ne}{c}\frac{dx}{dt} + \frac{1}{c}\frac{\partial X}{\partial t} = \frac{\partial\gamma}{\partial y} - \frac{\partial\beta}{\partial z}.$$

N is the number of charged particles per unit volume at the point in question and dx/dt is their average velocity in the x direction. In the case of transparent media these charged particles consist of electrons and ions vibrating about their positions of equilibrium. In the case of metals in addition to these stationary electrons and ions we have free electrons. The current along a wire consists simply of a procession of the free electrons along the wire. The equation of a free electron under the action of a light wave can be obtained from (30) simply by putting $f = 0$. The term for the effect of the free electrons on the index of refraction and coefficient of absorption can be obtained from the term for the effect of the stationary electrons simply by putting $f = 0$. Thus suppose N denotes the number of free electrons per unit volume, and that the region of the spectrum under consideration is so far from the free periods, that the influence of the latter on the index of refraction is constant, and on the coefficient of absorption can be neglected. Then

$$(\nu - i\kappa)^2 = n^2 + \frac{4\pi Ne^2}{-mg^2 + ihg}$$
$$= n^2 + \frac{2Ne^2\lambda}{c\left(-\frac{m2\pi c}{\lambda} + ih\right)},$$

where n^2 represents the effect of the stationary electrons and ions. This equation resolves into

$$\nu^2 - \kappa^2 = n^2 - \frac{4\pi Ne^2 m}{\left(\frac{m2\pi c}{\lambda}\right)^2 + h^2} \quad . \quad . \quad . \quad (44)$$

and

$$\nu\kappa = \frac{hNe^2\lambda}{c\left\{\left(\frac{m2\pi c}{\lambda}\right)^2 + h^2\right\}} \quad . \quad . \quad . \quad (45)$$

Now consider the equation of motion of the free electron

$$m\frac{d^2x}{dt^2} + h\frac{dx}{dt} = -eX.$$

Let us assume that X is constant. Then the equation can be written

$$\frac{d}{dt}\left(\frac{dx}{dt} + \frac{e}{h}X\right) + \frac{h}{m}\left(\frac{dx}{dt} + \frac{e}{h}X\right) = 0,$$

and has the solution

$$\frac{dx}{dt} + \frac{e}{h}X = e^{-\frac{h}{m}t}.$$

Hence when the steady state is reached the term on the right becomes 0 and

$$\frac{dx}{dt} = -\frac{e}{h}X.$$

The current per unit area in the x direction is $-Ne\,dx/dt$ or $(Ne^2/h)X$. Consequently the conductivity according to the usual definition is given by

$$\sigma = \frac{Ne^2}{h}.$$

If we return now to equation (45) and substitute σh for Ne^2, it becomes

$$\nu\kappa = \frac{\sigma h^2\lambda}{c\left\{\left(\frac{m2\pi c}{\lambda}\right)^2 + h^2\right\}} = \frac{\sigma\tau}{\left(\frac{2\pi m}{h\tau}\right)^2 + 1} \quad . \quad . \quad (46)$$

when τ is written for λ/c. We see at once why the relation $\nu\kappa = \sigma\tau$ is not fulfilled. It is on account of the term $(2\pi m)^2/(h\tau)^2$ in the denominator, which becomes zero only when τ is infinite. $\nu\kappa$ must be less than $\sigma\tau$. This is the case; for example in the case of mercury $\nu\kappa = 8{\cdot}6$ when $\sigma\tau = 20$.

According to (46) when $\nu\kappa$, σ, and τ are known h can be calculated. Hence N can be calculated by substitution in the expression for σ. It is found in this way that the number of free electrons is of the same magnitude as the number of molecules.

In the case of electrolytes σ is very much smaller than in the case of metals. Hence, since $\nu\kappa$ is always less than $\sigma\tau$, they do not absorb light appreciably.

If we consider equation (44) we see that κ can be greater than ν on account of the second term on the right being negative, more especially if h is small, which happens when the conductivity is great.

§ When λ is large equation (44) becomes

$$\nu^2 - \kappa^2 = n^2 - \frac{4\pi N e^2 m}{h^2},$$

i.e. the right-hand side is constant. But under the same circumstances $\nu\kappa = \sigma\tau$. Now σ is constant. Hence when τ is very large the product $\nu\kappa$ is very large, but the difference $\nu^2 - \kappa^2$ remains finite. Consequently ν must become approximately equal to κ.

The expression for the fraction of light reflected at perpendicular incidence by a metal mirror was given by

$$R = \frac{\nu^2 + \kappa^2 + 1 - 2\nu}{\nu^2 + \kappa^2 + 1 + 2\nu}.$$

Hence
$$1 - R = \frac{4\nu}{\nu^2 + \kappa^2 + 1 + 2\nu}.$$

If ν is put equal to κ this becomes

$$1 - R = \frac{2}{\kappa},$$

since κ is considerably greater than 1. Put $\nu = \kappa$ in $\nu\kappa = \sigma\tau$ and we obtain $\kappa = \sqrt{\sigma\tau}$. Hence

$$1 - R = \frac{2}{\sqrt{\sigma\tau}},$$

or, if R is measured in per cent,

$$100 - R = \frac{200}{\sqrt{\sigma\tau}} \qquad . \quad . \quad . \quad . \quad . \quad (47)$$

Rubens and Hagen measured the reflecting powers of various metals in the infra-red both by means of a reflecting spectrometer fitted with a fluorite prism and by the method of residual rays. They found that at the wave-length 12μ the above relation was satisfied. Thus at this wave-length the motion of the electrons has little influence and Maxwell's original theory is satisfied.

Equation (47), it should be noted, provides a method, although an inaccurate one, of comparing conductivities optically.

Theory of the Faraday Effect. If a piece of dense glass is placed in a strong magnetic field and a plane polarised beam of light sent through it in the direction of the lines of force, the plane of polarisation of the beam is rotated, the rotation being proportional to the strength of the field. This phenomenon has already been described in Chapter XIII and is known as the Faraday effect. It can be explained very readily in terms of the motion of the electrons contained inside the body.

Suppose for the sake of simplicity that there is only one kind of electron inside the body, in number N per unit volume, and that these electrons are vibrating about their positions of equilibrium. Neglect also the frictional resistance to the motion of the electrons, and suppose that the magnetic field is parallel to the Z axis and that the wave is propagated in this direction. Then, as may be seen from the theory of the Zeeman effect on p. 285, the equations of motion of the typical electron are

$$\frac{d^2x}{dt^2} + \frac{f}{m}x + \mathrm{H}\frac{e}{m}\frac{dy}{dt} = -\frac{e}{m}\mathrm{X},$$
$$\frac{d^2y}{dt^2} + \frac{f}{m}y - \mathrm{H}\frac{e}{m}\frac{dx}{dt} = -\frac{e}{m}\mathrm{Y},$$
$$\frac{d^2z}{dt^2} + \frac{f}{m}z = 0.$$

The difference in the sign of the H terms as compared with p. 286 is due to the fact that here the charge is $-e$. If we multiply the second of these equations by i and add it to the first, they combine into

$$\frac{d^2}{dt^2}(x + iy) + \frac{f}{m}(x + iy) - i\mathrm{H}\frac{e}{m}\frac{d}{dt}(x + iy) = -\frac{e}{m}(\mathrm{X} + i\mathrm{Y}) \quad . \quad (48)$$

According to (26) we have

$$-\frac{4\pi \mathrm{N}e}{c}\frac{dx}{dt} + \frac{1}{c}\frac{\partial \mathrm{X}}{\partial t} = \frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z}.$$

If we differentiate this equation with regard to t and substitute for β and γ from the second three equations of the electromagnetic field, it becomes on the assumption that $\mathrm{K} = \mu = 1$

$$-\frac{4\pi \mathrm{N}e}{c^2}\frac{d^2x}{dt^2} + \frac{1}{c^2}\frac{\partial^2 \mathrm{X}}{\partial t^2} = \frac{\partial^2 \mathrm{X}}{\partial x^2} + \frac{\partial^2 \mathrm{X}}{\partial y^2} + \frac{\partial^2 \mathrm{X}}{\partial z^2}.\text{*}$$

If we multiply the similar equation for Y by i and add it to this one, we obtain

$$-\frac{4\pi \mathrm{N}e}{c^2}\frac{d^2}{dt^2}(x + iy) + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}(\mathrm{X} + i\mathrm{Y}) = \frac{\partial^2}{\partial x^2}(\mathrm{X} + i\mathrm{Y}) + \frac{\partial^2}{\partial y^2}(\mathrm{X} + i\mathrm{Y})$$
$$+ \frac{\partial^2}{\partial z^2}(\mathrm{X} + i\mathrm{Y}) \quad . \quad . \quad . \quad . \quad (49)$$

This equation must be solved in conjunction with (48). Let us assume that t occurs in $x + iy$ and $\mathrm{X} + i\mathrm{Y}$ solely in the factor e^{igt}. Then (48) becomes

$$(x + iy)\left(-g^2 + \frac{f}{m} + \mathrm{H}\frac{e}{m}g\right) = -\frac{e}{m}(\mathrm{X} + i\mathrm{Y}),$$

* x has different meanings on the right-hand side and left-hand side of this equation. On the left-hand side it is the displacement of an electron, and on the right-hand side the coordinate of a point in space.

and on substituting for $x + iy$ in (49) we obtain

$$\frac{4\pi \mathrm{N}e}{c^2}\frac{1}{-g^2 + \frac{f}{m} + \mathrm{H}\frac{eg}{m}}\frac{e}{m}\frac{\partial^2}{\partial t^2}(\mathrm{X} + i\mathrm{Y}) + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}(\mathrm{X} + i\mathrm{Y})$$
$$= \frac{\partial^2}{\partial x^2}(\mathrm{X} + i\mathrm{Y}) + \frac{\partial^2}{\partial y^2}(\mathrm{X} + i\mathrm{Y}) + \frac{\partial^2}{\partial z^2}(\mathrm{X} + i\mathrm{Y}).$$

This represents the propagation of a wave in a medium, the square of the index of refraction of which is given by

$$n_1^2 = 1 + \frac{4\pi \mathrm{N}e^2}{-mg^2 + f + \mathrm{H}eg}.$$

The solution is therefore given by

$$\mathrm{X} + i\mathrm{Y} = e^{ig(t - zn_1/c)}$$

or
$$\mathrm{X} = \cos g\left(t - \frac{zn_1}{c}\right), \quad \mathrm{Y} = \sin g\left(t - \frac{zn_1}{c}\right).$$

As has already been shown on p. 221 this represents a left-handed circularly polarised wave.

Instead of finding equations for $x + iy$ and $\mathrm{X} + i\mathrm{Y}$, we might have taken $x - iy$ and $\mathrm{X} - i\mathrm{Y}$. The only difference in this case would be in the sign of the H term; hence, if we denote the index of refraction in this case by n_2,

$$n_2^2 = 1 + \frac{4\pi \mathrm{N}e^2}{-mg^2 + f - \mathrm{H}eg}.$$

The solution in this case is given by

$$\mathrm{X} = \cos g\left(t - \frac{zn_2}{c}\right), \quad \mathrm{Y} = -\sin g\left(t - \frac{zn_2}{c}\right),$$

and represents a right-handed circularly polarised wave travelling in the direction of the z axis.

In the medium in the magnetic field, then, X cannot occur apart from Y. Only circularly polarised waves can be propagated. If a plane polarised beam enters such a medium, it is decomposed into two circularly polarised waves which are propagated with different velocities c/n_1 and c/n_2. As was shown on p. 222 this is equivalent to a rotation of the original plane of polarisation at a rate of

$$\frac{g(n_2 - n_1)}{2c}$$

radians per cm. Now

$$n_2^2 - n_1^2 = 4\pi \mathrm{N}e^2\left(\frac{1}{-mg^2 + f - \mathrm{H}eg} - \frac{1}{-mg^2 + f + \mathrm{H}eg}\right)$$
$$= 4\pi \mathrm{N}e^2\frac{2\mathrm{H}eg}{(f - mg^2)^2 - (\mathrm{H}eg)^2} = 4\pi \mathrm{N}e^2\frac{2\mathrm{H}eg}{(f - mg^2)^2},$$

since $(\mathrm{H}eg)^2$ is small compared with the other term in the denominator. If we write $n_1 + n_2 = 2n$, where n is approximately the refractive index in the absence of a magnetic field,

$$n_2^2 - n_1^2 = (n_2 - n_1)2n,$$

and the rotation per cm. is given by

$$\frac{2\pi \mathsf{N} e^3 g^2 \mathsf{H}}{cn(f - mg^2)^2}.$$

If, as before, we substitute $m/f = (\lambda_0/2\pi c)^2$ and write $g = 2\pi c/\lambda$, this becomes

$$\frac{\mathsf{N} e^3 \lambda_0^4}{2m^2 \pi c^3 n} \frac{\lambda^2 \mathsf{H}}{(\lambda^2 - \lambda_0^2)^2} \quad . \quad . \quad . \quad . \quad . \quad (50)$$

Let us suppose that λ_0 is in the ultra-violet, then as λ diminishes, the magnitude of the rotation increases. Also the rotation is proportional to H and is greater if λ_0 is near the visible spectrum, i.e. if the medium has a high refractive index, all which is in accordance with experiment.

If we pass through a well-defined absorption band, according to (50) the rotation should increase to a very high value in the centre of the band and then decrease again. It has the same sign and is approximately symmetrical on both sides of the band. But formula (50) is only approximate. If the friction terms in the equation of the electron are not neglected it is found that the band splits into two components, the distance between which is proportional to the field strength, and that between the components the rotation is in the other direction. An absorption Zeeman effect takes place. The one component absorbs right-handed circularly polarised light and the other left-handed circularly polarised light. This absorption or "inverse" Zeeman effect, as it has been called, has been observed by J. Becquerel in the case of certain crystals containing didymium.

According to our investigation n_2 is greater than n_1, i.e. the right-handed wave is the slower. Consequently the rotation of the plane of polarisation is left-handed, or in the same direction as the current producing the field. Had the sign of e been different, the rotation would have been the other way. The fact, that in the great majority of substances the rotation is in the same direction as the current producing the field, shows that the vibrators are electrons.

The above theory of the Faraday effect was first given by W. Voigt.

Mechanical Analogy. The fact, that the anomalous dispersion of a medium could be explained by means of the forced vibrations of particles inside it, was first shown by Maxwell in 1869, but did not receive attention until put forward by Sellmeier two years later. Sellmeier used the elastic solid theory of light; he assumed that the medium was permeated by the ether and that the particles of the medium were attached to ether particles but not rigidly. Each material particle could execute vibrations about the ether particle to which it was attached. When the wave passed through the medium, the ether particles vibrated and the material particles attached to them executed forced vibrations.

In this section we shall calculate the velocity of a sine wave along a cord, which is stretched horizontally, and to each unit of length of which there is a number of little pendulums attached, each of mass m and length l. We shall suppose that the cord is displaced in a horizontal plane, that T is the stretching force in it and that ρ is its mass per unit length. This problem possesses a close analogy to Sellmeier's theory of dispersion, and also throws a great deal of light on the electromagnetic theory of dispersion.

During the passage of the wave each particle in the cord executes a S.H.M. in a horizontal direction at right angles to the undisturbed position of the cord. Since the point of support of each pendulum executes a S.H.M. in this direction, the pendulum itself executes vibrations in a vertical plane through this direction. At first the motion of the pendulum consists of two parts, the free vibration, the period of which is $2\pi\sqrt{l/g}$, and the forced vibration, which has the same period as the point of support. But the free vibration soon dies down and only the forced vibration is left. Let Aa and Bb be the two extreme positions of one of the pendulums when it is executing forced vibrations, a and b being the turning positions of the point of support. Produce Aa and Bb to meet at O. If the pendulum were a simple one vibrating about O, the motion of the part Aa would obviously be unaltered.

Now suppose that the pendulum is in the intermediate position Cc making an angle θ with the vertical. The stretching force in the cord is $mg\cos\theta$. The vertical component of this is $mg\cos^2\theta$, which may be taken equal to mg since θ is small; the horizontal component is $mg\sin\theta$. Let OC = L; L is the length of the equivalent simple pendulum which would have the

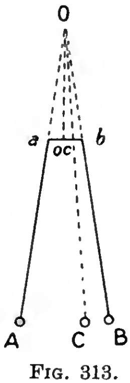


Fig. 313.

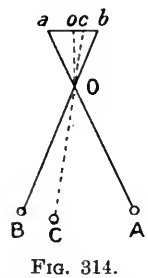


Fig. 314.

same period as the point of support has. Then Oc = CO − Cc = L − l and the horizontal component of the stretching force is

$$mg\frac{oc}{\mathrm{L}-l} = \frac{mg}{\mathrm{L}-l} \times \text{displacement of point of support.}$$

If L is smaller than l, the bob of the pendulum is in the opposite phase to the point of support, and we have fig. 314 instead of fig. 313, but the result is the same.

Let fig. 315 be a plan of a portion of the cord, and suppose that the wave is passing from left to right with velocity v. Consider the motion of an element of length FG with its midpoint at P, the displacement of which from its equilibrium position is NP. Let R be the radius of curvature at P. Then FG is acted on by two forces, the reaction of the pendulums attached to it and the resultant of the stretching forces at its two ends. The number of pendulums attached to the length FG is nFG, hence the first force is

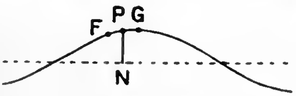


Fig. 315.

$$n\mathsf{FG}\,\frac{mg}{\mathsf{L}-l}\,\mathsf{NP}.$$

It may be shown by the triangle of forces that the second force is

$$\mathsf{FG}\,\frac{\mathsf{T}}{\mathsf{R}}.$$

To find the acceleration of the element suppose a velocity v from right to left superimposed on the whole diagram. This does not alter the acceleration of FG. The latter is then the acceleration of a particle moving round a curve of radius of curvature R with velocity v and is consequently v^2/R. Equating the rate of change of momentum of FG to the resultant force on it we find

$$\rho\mathsf{FG}\,\frac{v^2}{\mathsf{R}} = \mathsf{FG}\,\frac{\mathsf{T}}{\mathsf{R}} - n\mathsf{FG}\,\frac{mg}{\mathsf{L}-l}\,\mathsf{NP} \qquad . \quad . \quad . \quad (51)$$

It is shown in books on the Differential Calculus that the radius of curvature of a curve is given by $\frac{1}{\mathsf{R}} = \pm\frac{d^2y}{dx^2}$ when $\frac{dy}{dx}$ is small. Now the curve in fig. 315 may be represented by

$$y = e\sin\frac{2\pi x}{\lambda}.$$

Hence $\frac{1}{\mathsf{R}} = \pm\frac{d^2y}{dx^2} = \mp e\left(\frac{2\pi}{\lambda}\right)^2\sin\frac{2\pi x}{\lambda} = \mp\left(\frac{2\pi}{\lambda}\right)^2 y.$

On substituting $\mathsf{NP} = y = \mp\left(\frac{\lambda}{2\pi}\right)^2\frac{1}{\mathsf{R}}$ in (51) and simplifying, we obtain

$$v^2 = \frac{\mathsf{T}}{\rho} - \frac{nmg}{\rho(\mathsf{L}-l)}\left(\frac{\lambda}{2\pi}\right)^2.$$

When L is greater than l the ambiguous sign must be chosen so as to diminish the velocity. Write $\mathsf{T}/\rho = v_0^2$, since it is the square of the velocity when there are no pendulums attached, write τ for the period of the wave and τ_0 for the free period of the pendulum. Then

$$v^2 = v_0^2 - \frac{\mathsf{M}v_0^2\tau^2}{\tau^2-\tau_0^2}$$

where $\mathsf{M} = nmv^2/\mathsf{T}$. If we write μ for the ratio of the velocities, this gives

$$\mu^2 = 1 - \frac{\mathsf{M}\tau^2}{\tau^2-\tau_0^2},$$

an equation of the same form as (33). To represent a medium with two different classes of electrons we require a cord to which are attached two sets of pendulums with different lengths.

EXAMPLES.

(1) The indices of refraction of carbon bisulphide are given for seven different wave-lengths at 0° and 20° C. in the tables at the end of the book. Its coefficient of cubical expansion is represented by $a + 2bt$, where t is the temperature in degrees Centigrade, a = ·0011398, and b = ·00000137. Find whether the change in the refractive index is best represented by Gladstone and Dale's, Lorenz and Lorentz's, or Newton's formula.

(2) Show that the curve given in fig. 285, which represents Ives' results for the sensitiveness of the eye to light of different colours, can be represented fairly well by a formula similar to (37). Determine the most favourable values of the constants of the formula, and show by a graph how closely it fits the experimental results.

CHAPTER XXV.

THEORY OF RADIATION.

It has already been stated in Chapter XIV that in 1860 Kirchhoff published the law, that the ratio of the radiating power to the absorbing power of all bodies is the same and a function of the wave-length and the temperature. By the radiating power of a body is meant the quantity of heat radiated from unit area of its surface in unit time, and by the absorbing power is meant the fraction of the energy incident on the surface, that is absorbed by the body. This law has been the point of departure of some very interesting and far-reaching experimental and theoretical investigations.

Kirchhoff gave a rigorous theoretical proof of the law, and since his time there have been other proofs given. They are all, however, somewhat abstract, so we shall content ourselves with the proof of one simple case of the law. Kirchhoff gave this proof before he gave the rigorous proof.

Let CDFB, GJKH (fig. 316) be pieces of two bodies. They have the form of slabs extending to infinity on all sides. Let the faces CB, JK, which are turned away from one another, be impervious to heat. Let E_1, A_1 denote the radiating power and absorbing power of the body on the left, and E_2, A_2 the radiating power and absorbing power of the body on the right.

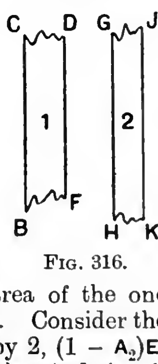


Fig. 316.

Suppose that there is equilibrium of temperature between the two bodies. Then according to Prévost's theory of exchanges each receives as much heat as it radiates. From the shape of the bodies it is evident that the radiation from each must be normal to their surfaces. It is sufficient, therefore, to consider unit area of the one surface and the unit area of the other surface opposite it. Consider the energy E_1 originally emitted from 1. A_2E_1 is absorbed by 2, $(1 - A_2)E_1$ reflected. Of the reflected energy $(1 - A_2)A_1E_1$ is absorbed by 1, $(1 - A_2)(1 - A_1)E_1$ reflected. Of this $(1 - A_2)(1 - A_1)A_2E_1$ is absorbed by 2, and so on. It is easy to see that the quantity absorbed by 1 is

$$(1 - A_2)A_1E_1 + (1 - A_2)^2(1 - A_1)A_1E_1 + (1 - A_2)^3(1 - A_1)^2A_1E_1 \ldots$$
$$= (1 - A_2)A_1E_1\{1 + (1 - A_1)(1 - A_2) + (1 - A_1)^2(1 - A_2)^2 \ldots\}$$
$$= \frac{(1 - A_2)A_1E_1}{1 - (1 - A_1)(1 - A_2)},$$

and the quantity absorbed by 2 is

$$A_2E_1 + (1 - A_1)(1 - A_2)A_2E_1 + (1 - A_1)^2(1 - A_2)^2A_2E_1 \dots$$
$$= \frac{A_2E_1}{1 - (1 - A_1)(1 - A_2)}.$$

If we interchange the suffixes we obtain the quantity of the energy E_2 that is absorbed by 1. It is

$$\frac{A_1E_2}{1 - (1 - A_1)(1 - A_2)}.$$

The total quantity of energy absorbed by 1 is

$$\frac{(1 - A_2)A_1E_1}{1 - (1 - A_1)(1 - A_2)} + \frac{A_1E_2}{1 - (1 - A_1)(1 - A_2)} = \frac{(1 - A_2)A_1E_1 + A_1E_2}{A_1 + A_2 - A_1A_2}.$$

This must equal E_1. Therefore

$$(1 - A_2)A_1E_1 + A_1E_2 = (A_1 + A_2 - A_1A_2)E_1.$$

Hence
$$A_1E_2 = A_2E_1$$
or
$$E_1/A_1 = E_2/A_2,$$
which proves the proposition.

§ The law has been tested experimentally in various cases and always found in agreement with facts. It is illustrated by many elementary experiments. For example, if letters are written with ink on a piece of bright platinum foil and the latter then heated in the bunsen flame, the writing stands out brighter than the surrounding foil. It absorbs more and consequently emits more.

The only form of energy considered by the proof is heat. Hence the law holds only for temperature radiation. Under this limitation it is universally accepted as true, being fulfilled in the case of all glowing solids and also in the case of some gases, such as carbon dioxide. Carbon dioxide when heated emits bands at $2{\cdot}8\mu$ and $4{\cdot}3\mu$. They are prominent in the spectrum of the bunsen burner. H. Schmidt measured the absorption and emission of these bands and found they obeyed Kirchhoff's law.

Band and line spectra cannot be produced as a result of heating alone, but only as the result of the liberation of chemical or electrical energy. Sources emitting bands and lines absorb these bands and lines, so the law is obeyed qualitatively in their case, but the ratio of E/A for these sources is not the same as for a glowing solid.

If in the expression for the law we write $A = 0$, in order that the ratio may remain finite E must also be put $= 0$. Now A may be zero for either of two reasons; the body may be perfectly transparent or its surface may reflect all the radiation that falls on it. We arrive therefore at the important result that a body cannot emit those rays for which it is transparent, or for which its surface acts as a perfect mirror.

In the rigorous proof of his law Kirchhoff introduced the idea of a perfectly black body, i.e. one for which $A = 1$. According to the table given on p. 362, A has the value ·988 for black cloth and the value ·996 for black velvet, hence neither of these substances is perfectly black.

Let S be the radiating power of a perfectly black body, and A and E the absorbing and radiating powers of any other body. Then

$$S = E/A.$$

Hence for any given temperature and wave-length no body can radiate more than a perfectly black body. S is, of course, a function of the wave-length and the temperature. It is sometimes called Kirchhoff's function. Kirchhoff pointed out the importance of determining it, because it enabled either E or A to be calculated when the other was known, and he stated that it would undoubtedly be of a simple form, since it was independent of the properties of any particular body.

§ It would be extremely difficult to measure E and A for one particular body for different wave-lengths and thus determine S, for, of course, A would have to be measured at the same temperature as E—that is, while the body is radiating. Paschen commenced a series of researches in 1892 with the purpose of determining S, the method being to measure E for a series of bodies which were more or less black. Then, if these bodies were arranged in order of their "blackness," it would be possible by a species of extrapolation to arrive at the behaviour of the ideal black body. Paschen used a piece of platinum foil folded double, in the fold of which a thermo-element was placed for the purpose of determining its temperature. The foil was heated by a current from a secondary battery, and was coated with the substances the radiation from which was to be examined. The radiation was measured with a bolometer: the spectral apparatus had concave mirrors instead of lenses, and the prism was of fluorite. The substances used were carbon filaments, platinum foil, platinum foil covered with iron oxide, copper oxide, and soot. They all behaved in much the same manner with the exception of platinum, which radiated considerably less than the others.

The next important step was the experimental realisation of the perfectly black body by Lummer and Pringsheim. A perfectly black body is one that absorbs all the rays that fall on it and reflects none. Consider the accompanying diagram (fig. 317). It represents a section of a hollow sphere which has a small opening at AB. R is a ray which enters the opening and is reflected in succession at C, D, E, and F. The inside surface of the sphere is blackened. Only a small portion of the energy is reflected each time, the greater part being absorbed. The hole AB is of such a size that the chances of the ray finding its way out again are very small. Light falling on AB is practically all absorbed. Consequently, if the sphere be maintained at a uniform temperature, sufficiently high to make its inner surface radiate out heat appreciably, the radiation from AB will be the radiation of a black body for that

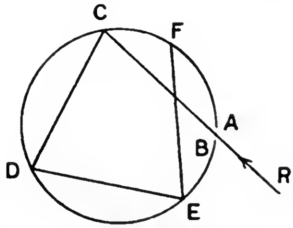


FIG. 317.

temperature, or black radiation, as it is called. Also, if any body be heated inside the sphere to the temperature of the latter, the radiation issuing from its surface through the opening will be black radiation.

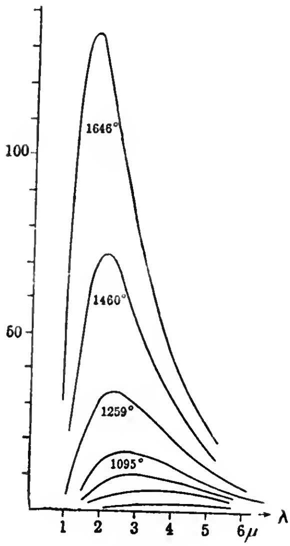


Fig. 318.

For low temperatures double-walled vessels were used, the space between the walls being filled with steam, ice, carbon dioxide snow, or liquid air, so as to keep the interior at a uniform temperature. The radiation escaped by means of a tube. For high temperatures an electrically heated body of porcelain was used.

With this apparatus Lummer and Pringsheim first proved that the radiation from a black body was proportional to the fourth power of the absolute temperature. This law had been enunciated by Stefan as a result of observations made on the rate of cooling of thermometers with blackened bulbs, etc. Then, in 1899, they published energy curves taken with the same radiators. Some of these energy curves are reproduced in the accompanying diagram (fig. 318). Thus Kirchhoff's function S was determined as a function of λ for several different values of T.

Density of Equilibrium Radiation in an Enclosed Vessel. At the same time as Paschen and Lummer and Pringsheim were seeking to measure S experimentally, endeavours were being made to determine S as a function of λ and T purely by theoretical reasoning. The results obtained have been very important, but as the reasoning leading up to them is long and abstract, only a short sketch of it will be given here.

The theoretical investigations are based on the idea of radiation being in equilibrium with matter. To understand this idea let us revert to fig. 316. At any time there is a flow of energy from 1 to 2 and a flow from 2 to 1. On the assumption that the velocity of the radiation is c let us calculate the density of the flow from 1 to 2.

The E_1 emitted by 1, if allowed to flow forward without obstruction, would fill a cylinder of length c and unit cross-sectional area. Hence it gives a density of E_1/c. The portion of E_2 emitted by 2 and reflected by 1 gives a density of $(1 - \mathsf{A}_1)\mathsf{E}_2/c$. The portion of E_1 reflected by both 2 and 1 gives $(1 - \mathsf{A}_1)(1 - \mathsf{A}_2)\mathsf{E}_1/c$. The portion of E_2 reflected by 1 and 2 and then again by 1 gives $(1 - \mathsf{A}_1)^2(1 - \mathsf{A}_2)\mathsf{E}_2/c$. If we proceed in this way, we find that the density of the resultant stream from 1 to 2 is given by

$$\frac{1}{c}\{E_1 + (1 - A_1)E_2 + (1 - A_1)(1 - A_2)E_1 + (1 - A_1)^2(1 - A_2)E_2 \ldots\}$$
$$= \frac{1}{c}\left\{\frac{E_1}{1-(1-A_1)(1-A_2)} + \frac{E_2(1 - A_1)}{1 - (1 - A_1)(1 - A_2)}\right\}$$
$$= \frac{1}{c(A_1 + A_2 - A_1A_2)}(E_1 + E_2 - A_1E_2).$$

If we substitute for E_2 from $A_1E_2 = A_2E_1$, the expression reduces to

$$\frac{E_1}{cA_1} \text{ or } \frac{S}{c}.$$

The density of the resultant stream in the other direction has, of course, the same value. Thus, when thermal equilibrium is established, the density of the radiation in the space enclosed between 1 and 2 is independent of the materials of which these substances are composed, and, if there is a small opening made in one of the slabs at P, the density of the radiation issuing from P is S/c. We thus see why the radiation from a hollow vessel is always black radiation, even though the inside of the vessel reflects quite well.

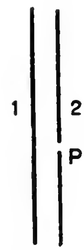


Fig. 319.

It is not necessary for the enclosed space to have the form shown in fig. 319 in order that the expression for the density may equal S/c; the enclosure may have any shape whatever. When a system is in thermal equilibrium the radiation in every direction for every wave-length has a definite density. If the surrounding bodies become colder, radiation is absorbed and the density becomes less; if they become warmer, radiation is emitted and the density in space becomes greater. We thus arrive at the idea of radiation in space being in equilibrium with matter at any temperature.

Boltzmann's Ether Engine. In the theory of the steam engine much use is made of a series of changes known as the Carnot cycle. A definite mass of gas can be put through a Carnot cycle by enclosing it in a cylinder with a tightly fitting piston as is shown in fig. 320, supplying a quantity of heat to it, and allowing it to expand isothermally. It is then allowed to expand further without gain or loss of heat. During this expansion its temperature falls. It is next allowed to contract isothermally, and during this operation a quantity of heat is taken from it. It is finally made to contract without gain or loss of heat, and during this operation the temperature increases to its initial value. The gas has then returned to its initial condition and the cycle of operations is completed.

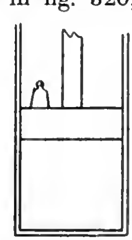


Fig. 320.

Suppose that instead of containing gas the cylinder is a vacuum, and that the inside surface of the cylinder and the lower surface of the piston are perfect reflectors. Let us suppose also that there is radiation inside the cylinder. This radiation will be reflected backwards and forward and never be absorbed. Its density throughout the cylinder is constant, and it can also be said to have the temperature of the matter with which it would be in equilibrium. The radiation passes in the ether, so that the energy in the cylinder may be said to be contained in the ether, just as in the former case it was contained in the gas. The analogy extends still further ; it was found by Bartolli as a deduction from the second law of thermodynamics, that the radiation exerts a pressure on the sides of the cylinder, just as in the former case a pressure was exerted by the gas. The radiation in the cylinder can consequently be put through a Carnot cycle just as the gas was, and as in this case the working substance in the cycle is the luminiferous ether, the arrangement of cylinder and piston for putting it through the cycle has been called an ether engine.

By considering such a cycle Boltzmann proved that the total radiation from a black body was proportional to the fourth power of its absolute temperature. This law had been stated previously but erroneously by Stefan as holding good for all bodies, and as the amended version was due to Boltzmann, it is very often referred to as the Stefan-Boltzmann law.

Wien's Displacement Law. Stefan's law applies only to the total radiation from a black body and gives no information as to how it is distributed through the spectrum. It is well known that when a piece of carbon is heated it sends out first a black heat, then, as its temperature is raised, a red heat,* and finally a white heat. With rise of temperature, therefore, the wave-length of the radiation diminishes. This has been known for a long time, and in 1847 Draper determined the temperature at which all bodies began to send out red light. He found it to be 525° C., and the statement that all bodies begin to emit red rays at this temperature was known for a long time as Draper's law. As a matter of fact bodies do not all begin to radiate at the same temperature. The " blacker " the body is, the lower is the temperature at which visible radiation starts. But Draper heated all his specimens in an iron tube closed at one end, and hence by his method of experimenting made them all to some extent black bodies.

As the temperature is raised, then, the wave-lengths of the radiation alter. This alteration has been investigated mathematically by Wien. Suppose we have a quantity of radiation of a given density enclosed in a cylinder with a piston, the inside surface of the enclosure being a perfect mirror, and suppose that the piston is pushed slowly down, then the radiation will be compressed and its density increased. It will consequently be in equilibrium with matter at a higher temperature. It has already been shown in the section dealing with the Doppler principle on p. 281, that if a beam is incident at an angle θ on a mirror which is moving with a velocity v in the direction of its

* If the observations are made very carefully with an extended object in a very dark room, owing to the Purkinje effect (Chapter XIX) it appears grey before red.

normal, then after reflection the wave-length of the beam is diminished in the ratio of 1 to $1 - 2(v/c)\cos\theta$. Now as the piston moves down, its surface acts as a moving mirror, the radiation in the cylinder is being continuously reflected by it and consequently its wave-length is being diminished. Arguing on these lines Wien showed, that if λ_m is the wave-length of the maximum of the energy curve at absolute temperature T, then $\lambda_m T$ is constant. He also showed how, when the energy curve is known for a given temperature, it can be constructed for any other temperature. This result is known as Wien's displacement law.

Radiation Formulæ. By assuming a particular and not at all plausible constitution for the radiating body, Wien obtained the formula

$$S_\lambda = \frac{c_1}{\lambda^5} e^{-c_2/\lambda T}$$

which is known as Wien's radiation law. Whenever S refers only to one particular wave-length we shall henceforth denote it by S_λ; then

$$S = \int_0^\infty S_\lambda \, d\lambda.$$

Owing to the definition of S_λ it is, of course, independent of the size of the hole or the distance of the latter from the slit of the spectroscope.

By assuming that all the "standing waves" which it was possible to form inside an enclosure possessed the same quantity of energy, and by dividing the spectrum into intervals and counting how many waves had their periods in each interval, Rayleigh derived the formula

$$S_\lambda = \frac{c_1 T}{c_2 \lambda^4}.$$

Rayleigh's formula does not represent the experimental results. Wien's formula was satisfied by Paschen's results, but did not fit Lummer and Pringsheim's data so well. Investigation showed that it was the formula, not Lummer and Pringsheim's results, that was at fault, and finally Planck proposed the formula

$$S_\lambda = \frac{c_1}{\lambda^5} \frac{1}{e^{c_2/\lambda T} - 1},$$

which fits all results within error of observation and which is at present almost universally accepted.

In deriving his formula Planck considers each wave-length separately. He assumes the existence of an enclosure containing a great number of Hertzian oscillators all radiating and absorbing the same wave-length, but possessing at any given instant different quantities of energy. This distribution of the energy among the different oscillators occurs according to the laws of probability, and by using a very general definition of temperature the temperature of the system can be derived from the distribution of energy. Then by calculating the density of the radiation in the enclosure the value of S_λ for that particular value of λ can be found as a function of T.

In order to obtain the correct result Planck found it necessary to assume, to put the matter very crudely, that the oscillators did not radiate continuously, but emitted the energy in jerks or "quanta". This theory of quanta is, of course, opposed to all previous notions. It seems to find a confirmation in one or two other branches of physics, and has aroused a great deal of attention at present. But if we consider the modern view as to the nature of white light, which has been expounded in Chapter XXI, according to which white light consists of pulses and the harmonic waves into which it can be resolved are mere mathematical fictions having no physical reality, it does not seem justifiable to consider each different wave-length separately. If this is the case, Planck's whole argument falls to the ground.

Discussion of the Radiation Formula. It will be noticed that when λT is made very large, Planck's formula agrees with Rayleigh's, and when λT is made very small, it agrees with Wien's. In far the greater number of cases, indeed except when dealing with residual rays, Wien's formula does not differ much from Planck's, and, since it has the advantage of simplicity, it can be used as an approximation for the latter. From it a number of important results about black body radiation can be deduced.

We have
$$S_\lambda = c_1\lambda^{-5} e^{-c_2/\lambda T}.$$
To find for what value of λ the expression has a maximum take the logarithm of S_λ, differentiate it with respect to λ, and equate the result to zero.

This gives
$$\log S_\lambda = \log c_1 - 5 \log \lambda - \frac{c_2}{\lambda T},$$
$$-\frac{5}{\lambda} + \frac{c_2}{\lambda^2 T} = 0,$$
and finally
$$\lambda T = \frac{c_2}{5}.$$

Denote the value of λ corresponding to the maximum value of S_λ by λ_m. When T is measured in degrees absolute and λ in 10^{-4} cm., $c_2 = 14{,}500$. Therefore
$$\lambda_m T = 2900.$$
As the temperature increases, the maximum moves towards the visible spectrum. By substituting λ_m in the expression for S_λ, we find that the maximum value of S_λ increases as the fifth power of the absolute temperature. S_λ always increases with the temperature for every value of λ, no matter on which side of the maximum it is.

To find the total radiation, integrate $c_1\lambda^{-5} e^{-c_2/\lambda T}$ from $\lambda = 0$ to $\lambda = \infty$. The integration follows easily if λT is put $= 1/\theta$ and θ is regarded as the variable. The result is $6c_1 T^4/c_2^4$. This is Stefan's law.

The temperature radiation of bodies that are not black follows more complicated laws. The energy curves vary all the way from shapes not differing widely from that for the black body to curves with two or more sharp maxima. The wave-length of maximum intensity, however, always decreases as the temperature increases, and many substances, which radiate selectively at low temperatures, become similar to the black body at high temperatures. Of course, no body can radiate more than a black body for any wave-length and temperature.

The metals form an interesting class, and an attempt has been made to represent their energy curves by formulæ of the same form as Wien's, but with a different power of λ in place of the λ^{-5}. At ordinary temperatures they all reflect much better in the infra-red than in the visible spectrum. Hence their absorbing power and consequently their radiating power is less in the infra-red. This selective reflection seems to persist at high temperatures, and the energy curves resemble somewhat the curves for the black body, but with the ordinates on the infra-red side of the maximum diminished, and consequently the radiant efficiency increased. Some of Lummer and Pringsheim's curves representing the radiation from glowing platinum at different temperatures are given in fig. 321. The total radiation from platinum varies as the fifth power of the absolute temperature.

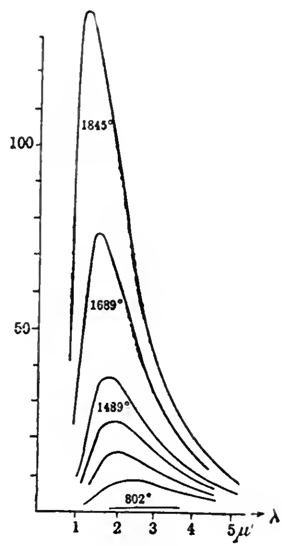


FIG. 321.

Optical Pyrometry. The theory of the black body has been applied in three different ways to the measurement of temperature. Temperature may be read by the hydrogen thermometer up to 600° C., and thence by the nitrogen thermometer up to 1150° C. Beyond this it is customary to use thermo-couples, which have been calibrated as far as the gas thermometers go, and to extrapolate for values beyond. Thermo-couples are, however, unsuitable for some purposes, and in any case they cannot be used beyond the melting-point of platinum. This is where optical pyrometry finds its use.

The first optical method consists in finding the position of λ_m and applying the equation $\lambda_m T = 2900$. The second consists in measuring the total radiation by some instrument such as Féry's thermo-electric telescope and using Stefan's law. In this instrument an image of the source is formed on a thermo-couple by a concave mirror, and the instrument is calibrated by pointing it at two sources of known temperature. The third method consists in measuring the brightness of the source for one wave-length in the visible spectrum, and calibrating the photometer by means of two sources of known temperature.

These methods give relative measurements satisfactorily and enable one to reproduce any temperature with accuracy, which is all that is required for industrial purposes; but they must be used with considerable care if absolute measurements are required. Many people in possession of optical pyrometers are unaware of the limitations of their instrument, and have too profound a faith in its empirical formula.

All three methods hold in the first instance only for the black body. The first method holds also for a "grey" body—that is one that reflects all colours equally well. In the case of a grey body the ordinates of the energy curve are supposed to be diminished in the same ratio all over, and the position of the maximum is unaltered. The first method is, however, not so accurate as the other two, for the position of the maximum cannot be easily determined. The last two methods if applied indiscriminately will not give the true temperature of a body but its "black" temperature—that is, the temperature of a black body which would have the same total radiation, or which would have the same brightness for the wave-length in question.

To get the true temperature we may proceed in two ways. On the first method, which is due to Lummer and Pringsheim, the scale of the instrument is calibrated by using glowing platinum as well as the black body. We thus have two determinations of the temperature, the second value being that which the body would have if it had the same properties as platinum. As platinum differs widely from the black body as a radiator, there is a presumption that the true temperature of the body will lie between the limits thus obtained. On the second method the body is placed inside a sphere, the inner surface of which is a good reflector, and the radiation through a hole in the surface examined. The radiation from the body is thus increased, until its black temperature equals its true temperature.

Optical pyrometry has been applied to the determination of the temperature of the sun, and values somewhat below 6000° C. have been obtained. If the sun is a black radiator, this is not far from its most economic temperature, i.e. the temperature at which the greatest proportion of its radiation falls within the limits of the visible spectrum. If the temperature were greater than this, too large a proportion of the radiation would fall in the ultra-violet. Through long ages of evolution our eyes have adapted themselves to the sun, and become most sensitive to these wave-lengths which it radiates most intensely.

In the ordinary gas or candle flame, the carbon glow lamp, and the crater of the carbon arc, the source of light is in each case glowing carbon, and the energy spectrum is approximately that of the black body. The fundamental difference is that of temperature. By means of the thermojunction the temperature of the luminous region of the gas flame has been found to be not more than 1780° C., and by optical methods the temperatures of the carbon glow lamp and arc crater have been found respectively to be below 1800° C. and between 3500° and 4000° C.

The Pressure of Light.* It has been shown from Maxwell's equations that there is momentum in an electromagnetic field as well as energy, that the direction of this momentum is the same as the direction of propagation of the energy, and that its value per unit volume is numerically equal to the energy per unit volume divided by the velocity of light. A pencil of light is consequently a stream of momentum.

Thus if a light wave is incident on a plane absorbing surface at an angle ϕ and if the energy falling per second per square centimetre held normal to the rays is E, the momentum received per second per square centimetre is $(\mathsf{E} \cos \phi)/c$. This produces

$$\text{Normal pressure} = \frac{\mathsf{E} \cos^2 \phi}{c}.$$

$$\text{Tangential stress} = \frac{\mathsf{E} \cos \phi \sin \phi}{c}.$$

If the wave is entirely absorbed both these forces exist.

If the stream is entirely reflected, the reflected pencil exerts an equal normal force and an equal and opposite tangential force and we have only the normal pressure of amount $2\mathsf{E} (\cos \phi)^2/c$.

If only a fraction r is reflected, the incident and reflected waves give

$$\text{Normal pressure} = \frac{(1 + r)\mathsf{E} \cos^2 \phi}{c}.$$

$$\text{Tangential stress} = \frac{(1 - r)\mathsf{E} \cos \phi \sin \phi}{c}.$$

A beam of light should exert a pressure, then, on a surface on which it falls. This pressure is, however, extremely small. For solar radiation at the earth's surface, when the absorption of the atmosphere is allowed for, $\mathsf{E} = 0{\cdot}175\ 10^7$ ergs/cm.^{2} sec. Consequently if the surface is a black one and the incidence is normal, the magnitude of the normal component is only

$$\frac{0{\cdot}175\ 10^7}{3\ 10^{10}} = 5{\cdot}8\ 10^{-5} \text{ dynes/sq. cm.}$$

A pressure was to be expected according to Newton's corpuscular theory, owing to the destruction of the momentum of the stream of particles which constituted the light ray according to that theory, and the pressure could be calculated on the corpuscular theory according to the ordinary laws of dynamics. An unsuccessful attempt was made to detect its existence experimentally as early as 1754. Unsuccessful attempts were made to detect the pressure by other physicists, that of Crookes, for example, leading to the discovery of the radiometer. An expression was derived for the pressure on the electromagnetic theory by Clerk Maxwell, and shortly afterwards it was shown independently by Bartolli from thermodynamical considerations, as has already been mentioned, that a light wave exerts a pressure on a surface on which it

* The matter of this and the two following sections is taken from a paper by Prof. J. H. Poynting, "Phil. Trans.," A., vol. 202, 1904, p. 538.

falls. The above method of deriving the formulæ is considerably simpler than Clerk Maxwell's, but the legitimacy of using the conception of optical momentum would not be universally admitted.

The existence of the normal component of the pressure was proved experimentally in 1900 by Lebedew, and independently immediately afterwards by Nichols and Hull. The method of experimenting was approximately the same in each case. Light from an arc lamp was concentrated on a suspended vane, made of thin platinum discs in Lebedew's arrangement and of thin silvered glass in Nichols and Hull's arrangement. The light produced a rotation of the vane and this rotation was read by a mirror attached to it. The most suitable pressure of air was sought in each case. The chief difficulty was the radiometer or gas action; the side on which the light falls gets heated, the pressure on that side rises and this rise in pressure produces a spurious effect. Lebedew eliminated the gas action by letting the light fall rapidly in succession on the two sides of the disc and taking the mean of the two deflections. The results in both cases were found to agree with theory.

The tangential stress was detected and measured by Poynting and Barlow in 1904. The angle of incidence used was 45°, and the magnitude of the result was found to agree with theory.

Pressure of Radiation on its Source. If rays of light are streams of momentum, it follows, since the momentum of rays and source in any direction is constant, that the rays must exert a backward pressure on their source.

If 1 sq. cm. of surface is emitting altogether energy R per second, and if $\mathsf{N}d\omega$ is the energy it is emitting through a cone $d\omega$ with axis along the normal, then in direction θ it is emitting $\mathsf{N}\ \cos\theta\ d\omega$ through a cone $d\omega$. Putting $d\omega = 2\pi\ \sin\theta\ d\theta$ and integrating over the hemisphere, we have

$$\mathsf{R} = \int_0^{\frac{\pi}{2}} \mathsf{N}\cos\theta\ .\ 2\pi\ \sin\theta\ d\theta = \pi\mathsf{N}.$$

If we draw a hemisphere, radius r, round the source as centre, the energy falling on area $r^2d\omega$ is $\mathsf{N}\ \cos\theta\ d\omega$ per second, and, since the velocity is c, the energy density just outside the surface on which it falls is $\mathsf{N}\ (\cos\theta)/(cr^2)$, and this is the rate at which momentum is being received, that is, it is the normal pressure. The total force on area $r^2d\omega$ is $\mathsf{N}\ \cos\theta\ d\omega/c$. This is the momentum sent out by the radiating square centimetre per second through the pencil with angle $d\omega$, in the direction θ, and is therefore the force on the square centimetre due to that pencil.

Resolving along the normal and in the surface we have

$$\text{Normal pressure} = \frac{\mathsf{N}\cos^2\theta d\omega}{c},$$

$$\text{Tangential stress} = \frac{\mathsf{N}\cos\theta\sin\theta d\omega}{c}.$$

Putting $d\omega = 2\pi \sin\theta\, d\theta$ and integrating over the hemisphere we get

Resultant normal thrust

$$= \int_0^{\frac{\pi}{2}} (\mathrm{N} \cos^2\theta\ 2\pi \sin\theta\ d\theta/c) = 2\pi\mathrm{N}/(3c) = 2\mathrm{R}/(3c).$$

Total tangential stress = 0, since the radiation is symmetrical about the normal.

Comparison of Radiation Pressure with Gravitational Attraction exerted by the Sun on a Small Body. Let the small body be a sphere of radius a and density ρ, let its surface be a perfect absorber, and let it be at one temperature throughout. Let us assume that it is exposed to solar radiation of intensity E ergs/sq. cm. sec. Then it receives a momentum

$$\frac{\pi a^2 \mathrm{E}}{c}$$

per second from the sun. Its own radiation being equal in all directions has zero resultant thrust.

The acceleration of gravity towards the sun at the distance of the earth is about 0·59 cm./sec. Thus we have

$$\frac{\text{Radiation pressure}}{\text{Gravitation pull}} = \frac{\pi a^2 \mathrm{E}}{c \times \frac{4}{3}\pi a^3 \rho \times 0{\cdot}59}.$$

The two will be equal when

$$a = \tfrac{3}{4}\,\frac{\mathrm{E}}{c\rho \times 0{\cdot}59}.$$

If we put $\rho = 1,\ \mathrm{E} = 0{\cdot}175\ 10^7,\ c = 3\ 10^{10}$,

we get $a = 74\ 10^{-6}$,

i.e. a body of diameter about two wave-lengths of red light would be equally attracted and repelled, if we could assume that a surface so small still continues to absorb. But when we are getting to dimensions comparable with a wave-length that assumption can no longer be made.

EXAMPLES.

(1) Show that, if Wien's formula is assumed for the black body, the maximum value of S_λ varies as T^5.

(2) Derive Stefan's law from Wien's radiation formula, giving in full all the steps of the integration. Show in the same way that Stefan's law can be derived from Planck's radiation formula.

(3) Show clearly that if $\lambda\mathrm{T}$ is made very large Planck's radiation formula reduces to Rayleigh's.

(4) Show that when $1/\theta$ is substituted for $\lambda\mathrm{T}$ in Wien's radiation formula and the integral taken with regard to the wave-length, the result is

$$c_1 \mathrm{T}^4 \int \theta^3 e^{-c_2\theta}\, d\theta.$$

If the curve $\theta^3 e^{-c_2\theta}$ is plotted and ordinates set up at values of θ corresponding to the ends of the visible spectrum for a definite temperature, show how the ratio of the area included between these ordinates to the area of the whole curve gives the radiant efficiency (cf. p. 353) for that temperature. Hence find the radiant efficiency of a black body at a temperature of 2000° abs.

(5) The total radiation from platinum varies as the fifth power of the absolute temperature. Show that the formula

$$S_\lambda = c_1\lambda^{-6} e^{-c_2/\lambda T}$$

satisfies this requirement. Show also, as is required for all metals, that this formula gives relatively less radiation in the infra-red than is done by the corresponding formula for the black body.

(6) Let E_1 denote the radiating power of an arc lamp for sodium light, and let E_2 denote the radiating power for the same wave-length of a sodium flame placed in front of it. Let a denote the fraction of the incident sodium light absorbed by the flame. The sides of the sodium flame are plane and at right angles to the line of vision of an observer. Show that the D lines will appear reversed if $E_1 a$ is greater than E_2.

(7) Show that an infinitely thick flame radiates like a black body. [Divide the flame into n parallel layers of equal thickness; let E be the radiating power of a single layer, and let a be the fraction of light absorbed by a layer. Then the radiating power of the whole flame is

$$E + E(1 - a) + E(1 - a)^2 + \ldots]$$

(8) It is found that, when a thermopile is exposed to the full radiation of the sun, the deflection on the scale is 81 times as great as when it is exposed to the radiation from a disc of platinum which subtends the same angle at the thermopile as the sun does. The "black" temperature of the platinum is 2000° abs., i.e. its surface is as bright as the surface of a black body at that temperature. Find the "black" temperature of the sun.

(9) The sun is 600,000 times as bright as the full moon. The brightness of the sky at full moon is 10^{-6} times the brightness of the moon's surface. Assuming that the total radiation from the sun and from the sky is proportional in each case to the visible radiation, show that the temperature of a small portion of matter in interstellar space cannot be greater than 7° abs.

CHAPTER XXVI.

THE RELATIVE MOTION OF MATTER AND ETHER.

Astronomical Aberration. We come now to discuss the nature of the ether. If light is a wave-motion propagated out from the sun, we must assume the existence of a medium for it to travel in, and this medium, which fills interstellar space and penetrates between the molecules of matter, is known as the luminiferous ether. It has usually been discussed fully at an early stage in books on light and attention drawn to its contradictory properties, how it opposes no measurable resistance to the motion of the planets through space, and yet acts as a rigid solid as far as the vibrations of light are concerned. But in this book the discussion of it has hitherto been evaded, because there is some doubt at present as to whether it really exists.

Fresnel regarded the ether as continuous and having a density and elasticity in the same way as ordinary matter has, only its density was infinitesimal compared with the density of ordinary matter. The density of the ether varied from medium to medium, being greater inside an optically dense medium such as glass than in air, but its elasticity was the same in all media. The velocity of light in any medium was inversely proportional to the square root of the density of the ether inside that medium. We have already seen that on this basis Fresnel was able to give a satisfactory theory of reflection and refraction.

Now what happens when matter moves through the ether? Does the ether contained in it move with it, or does the matter move through the ether like a wire framework through air?

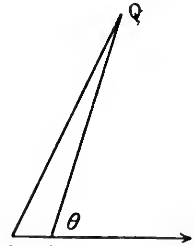


FIG. 322.

In 1726 the astronomer Bradley observed a phenomenon, known as the aberration of light, which apparently enables us to answer this question. It has already been described in Chapter VIII, and consists in an apparent displacement of the stars in the sky owing to the motion of the earth in its orbit about the sun. Suppose that an observer is stationed at O and looks at a star, the true direction of which is OQ, but that the earth is moving with velocity v in the direction OP. Let V be the velocity of light. In estimating v only the velocity of the earth in its orbit need be considered, the component due to the diurnal rotation being neglected. If the axis of the telescope were pointed in the direction OQ, owing to the lateral motion of the telescope the rays from the star would not travel down its axis. In order that this may take place the telescope must be pointed in the direction O_1Q, where $O_1O/OQ = v/V$. For if we suppose the centre of the object glass to be at Q, the cross-wires move from O_1 to O while the ray is travelling from Q to O, and consequently only if the telescope is pointed in the direction of O_1Q will the rays come to a focus on the cross-wires.

The phenomenon is exactly analogous to the case of a ship which is moving at a high speed, and which is being struck by shots fired at right angles to its line of motion. Owing to the motion of the ship the track of the shot through it will not be at right angles to its length, but the hole in the side at which the shot leaves will be further astern than the hole at which it enters. Thus the apparent direction of the cannon as estimated from the shot holes is ahead of the true direction, in the same way as the apparent direction of the star is ahead of its true direction.

The angle between the true and apparent directions of the star, namely, OQO_1, is given by

$$\frac{\sin OQO_1}{\sin QOO_1} = \frac{OO_1}{O_1Q}.$$

Since O_1O is small, we may write OQ for O_1Q and $\angle OQO_1$ for its sine. Hence the aberration

$$\angle OQO_1 = \frac{OO_1}{OQ} \sin QOO_1.$$

The maximum value of OQO_1 is obtained by putting $\sin\theta = 1$, and equals 20·47 seconds. As has already been stated, the value of the velocity of light obtained from measurements of the aberration agrees with the values obtained by the other methods.

In the above explanation we have assumed that the ether remains at rest while the telescope and the air inside it move through it. We can thus apparently assume that the ether remains always absolutely at rest.

Airy's Experiment. This assumption is, however, upset by an experiment due to Airy. He filled a telescope with water. The velocity of the ray down the telescope would consequently be diminished, while the velocity of the earth through space remains the same. Also the ray would be refracted on passing out of the object glass into the water. Now according to the foregoing theory on account of both these reasons the aberration should be greater. But it remains exactly the same.

We must therefore modify our assumption, and assume that the ether inside the telescope drifts with it when it is filled with water, but not when it is filled with air. The magnitude of this drift can easily be calculated from geometrical considerations.

For in fig. 323 let OQ represent the true direction and O_1Q the apparent direction of the star. When the rays enter the water at Q they are refracted. Now QO gives their direction of incidence and QO_1 the normal; consequently the direction of refraction is given by QS, where

$$\sin O_1QO = \mu \sin O_1QS.$$

Since the angles are small we may write this equation

$$O_1QO = \mu\, O_1QS$$

or simply

$$O_1O = \mu\, O_1S.$$

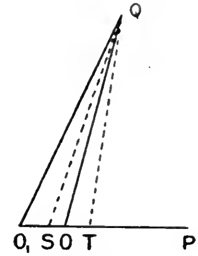


FIG. 323.

The velocity of the rays down the telescope is diminished in the ratio of μ to 1, consequently the time taken by them in passing down the telescope is increased in the ratio of 1 to μ and during this time the cross-wires would arrive at T instead of O, where $O_1T = \mu O_1O$.

We have to explain, therefore, how it is that the ray arrives at T instead of S, and this can be done only by assuming that the ether drifts the distance ST while the rays are passing down the tube. While the water moves the distance O_1T, the ether moves ST, or, in other words, the ether contained inside the water is moving in the same direction as the latter, but with a velocity only ST/O_1T of the velocity of the latter.

We have therefore for the velocity of the ether drift

$$\frac{ST}{O_1T}v = \frac{O_1T - O_1S}{O_1T}v = \left(1 - \frac{O_1S}{O_1T}\right)v$$
$$= \left(1 - \frac{O_1O/\mu}{\mu O_1O}\right)v = \left(1 - \frac{1}{\mu^2}\right)v.$$

This expression was first obtained by Fresnel. It becomes zero for air, i.e. when $\mu = 1$.

Fresnel's Method of Obtaining the Expression for the Ether Drift. Suppose that a glass plate is moving through the ether with velocity v, that the density of the ether inside the glass is ρ', and that its density in vacuo is ρ; ρ' is, of course, greater than ρ. Then it is clear that the ether inside the glass must be to a certain extent dragged with it, for if it remained at rest the glass would move away from the place where the ether had the greater density.

Let v' be the velocity of drift of the ether. Then since there is no flow round the edges of the plate, the quantity entering the front surface is ρv per unit area. The quantity leaving the back surface of the plate is $\rho'(v - v')$ per unit area. Since the quantity inside the plate is constant,

$$\rho v = \rho'(v - v') \quad \text{or} \quad v' = v\left(1 - \frac{\rho}{\rho'}\right).$$

But $\rho'/\rho = \mu^2$, the square of the index of refraction of the glass. Consequently on substituting we find

$$v' = v\left(1 - \frac{1}{\mu^2}\right),$$

the same result as before.

Fizeau's Experiment. Fresnel's formula for the ether drift was verified in a celebrated experiment carried out by Fizeau in 1859.

Fig. 324 shows the arrangement. As source a narrow slit at S illuminated by monochromatic light is taken. The rays from S are reflected by a glass plate G, made parallel by a lens L, and then pass through two apertures A and B. CD is a vessel which is divided into two parts by a partition, and through which water can be forced in the direction of the arrows. The pencils from the apertures enter and leave the vessel through parallel plates of glass. They are then received by a lens L′ and reflected by a mirror M at its focus, after which they pass back through L′, CD, L, and G to S′, but with their paths interchanged. The pencil which enters CD through A leaves it through B. The two pencils form interference bands at S′.

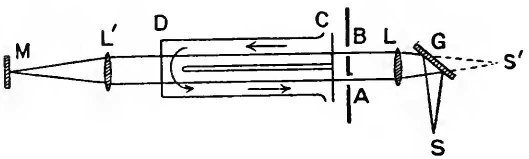


FIG. 324.

If the water is at rest, the paths of the two pencils are exactly similar. If, however, the water is being forced through, the pencil which enters through A travels against the current both going and returning. Hence, if according to Fresnel's theory the ether is dragged with the current, this pencil is retarded, while the other pencil which in both cases travels with the current is accelerated. The interference bands at S′ should therefore be displaced.

The phase difference introduced by the motion can easily be calculated. For let V be the velocity of light in vacuo, v the velocity of the water and μ its index of refraction, and let l be the length of the vessel CD. Then the difference of the times required by the two paths is

$$\frac{2l}{\frac{V}{\mu} - \left(1 - \frac{1}{\mu^2}\right)v} - \frac{2l}{\frac{V}{\mu} + \left(1 - \frac{1}{\mu^2}\right)v}$$

$$= \frac{4l\left(1 - \frac{1}{\mu^2}\right)v}{\frac{V^2}{\mu^2} - \left(1 - \frac{1}{\mu^2}\right)^2 v^2}.$$

From the difference of times in emerging from A and B and the velocity of light in air the path difference at S′ can be calculated.

Fizeau obtained a measurable result when the velocity of the water was 7 metres per second, and the result supported Fresnel's theory.

The experiment was repeated by Michelson and Morley in 1886 in a slightly different manner and with greater accuracy. Their results also bear out the foregoing theory.

Michelson and Morley's Experiment. Fizeau's experiment is unique in that it gives a positive effect. Many experiments have been made with the intention of detecting a possible effect of the earth's motion on optical phenomena, but the results have been wholly negative. The most celebrated of these experiments has been that made in 1887 by Michelson and Morley with Michelson's interferometer.

Fig. 325 represents the interferometer. Its use has already been

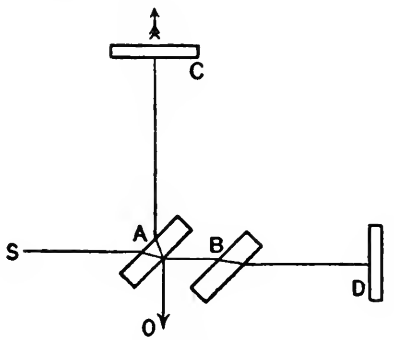


FIG. 325.

described on pp. 149 and 291. The light from the slit enters the glass plate A and divides into two beams at its second surface, which is half-silvered. One of these beams then goes back through the plate and is reflected back on its path by the mirror at C, while the other passes through the equally thick plate B and is reflected back on its path by the mirror at D. The two beams then superimpose again at the point where they separated, and interference bands are produced. These interference bands are observed by a telescope which is pointed in the direction OA. Let us suppose that the two paths are optically equal, and that the optical distance from the point of separation of the beams to either of the mirrors is d.

Let us suppose now that the earth is moving through space in the direction parallel to SD with a velocity v. Then the time taken by the light to go from A to D and back is

$$\frac{d}{V + v} + \frac{d}{V - v},$$

where V is the velocity of light. The equivalent optical length of this path is consequently

$$Vd\left(\frac{1}{V + v} + \frac{1}{V - v}\right) = \frac{2dV^2}{V^2 - v^2} = 2d\left(1 + \frac{v^2}{V^2}\right)$$

approximately. We have supposed in this calculation that the ether is at rest in the glass as well as in the air, but owing to the fraction of the path in the glass being very small this assumption does not affect the result appreciably.

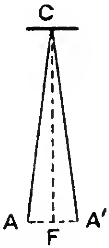


FIG. 326.

Let us calculate now the equivalent length of the path AC. Owing to the motion of the glass plate A in space from A to A′ (fig. 326), while the light is going from the plate to the mirror and back, the rays are no longer incident on C perpendicularly but they fall on it at an angle as shown in the figure. $\mathrm{CF} = d$ and $\mathrm{AF}/\mathrm{CF} = v/\mathrm{V}$. Hence the equivalent length is

$$\mathrm{AC} + \mathrm{CA}' = 2\mathrm{AC} = 2\sqrt{\mathrm{CF}^2 + \mathrm{AF}^2}$$

$$= 2\mathrm{CF}\left(1 + \frac{\mathrm{AF}^2}{\mathrm{CF}^2}\right)^{\frac{1}{2}} = 2\mathrm{CF} + \frac{\mathrm{AF}^2}{\mathrm{CF}} \text{ approximately}$$

$$= 2d + \frac{v^2 d}{\mathrm{V}^2} = 2d\left(1 + \frac{v^2}{2\mathrm{V}^2}\right).$$

The difference between the equivalent lengths of the two paths is consequently

$$\frac{dv^2}{\mathrm{V}^2}.$$

If the earth were to stop suddenly, then the interference bands would be displaced in the field.

It is impossible to start or stop the earth's motion suddenly, but the same effect can be obtained by rotating the whole apparatus through 90°, so that the path AC becomes parallel to the direction of the earth's motion and the path AD at right angles to it. This makes the difference the other way, so that the rotation produces a difference of length equal to $(2dv^2)/\mathrm{V}^2$.

In the first experiments tried the calculated shift was too small, so by means of multiple reflections d was increased to 11 metres. We have $v^2/\mathrm{V}^2 = 10^{-8}$. Hence the difference should have been $2{\cdot}2 \times 10^{-5}$ cms., or less than half a wave-length of sodium light. The actual displacement observed was certainly less than $\frac{1}{20}$ and probably less than $\frac{1}{40}$ of this amount. To avoid vibration and distortion of the apparatus due to rotating it, it was mounted on a heavy slab of stone floating in an iron trough, which was cemented in a low brick pier. This slab was kept in a slow rotation, the time of a complete revolution being about six minutes, and, in order to avoid the strains set up by stopping it or starting it, the observations were made while it was moving.

The apparent result of Michelson and Morley's experiment is to prove that the ether in the vicinity of the earth moves with the same velocity that the earth does, a result which is at variance with Fresnel's law of drift and the phenomena of astronomical aberration. In order to remove this contradiction Fitzgerald and Lorentz put forward the suggestion simultaneously, that the lineal dimensions of matter depend to a slight extent on its absolute motion through space. If the arm AD (fig. 325) which lies in the direction of the earth's motion is always $(dv^2)/(2V^2)$ shorter than the other, then the two paths would have the same equivalent length.

This suggestion is not so artificial as it seems at first sight, because in the first place the change required in the length is so very small. Since v^2/V^2 is only 10^{-8}, the diameter of the earth which is parallel to the direction of motion would be diminished only $2\frac{1}{2}$ inches by the motion. Also the length of a body depends on the forces between its molecules. These molecular forces may very well be electrical in their origin. If so, they would be transmitted by the ether and would be influenced by the motion of the molecules through the ether.

Sir Oliver Lodge made an experiment, in which two large steel discs 3 feet in diameter were mounted on a common axle, one above the other with their planes 1 inch apart, and rotated at a very high speed. Two interfering beams were passed round the space between the discs in opposite directions by means of a system of mirrors. Fig. 327 shows a plan of the arrangement. The two beams are separated and united again at the half-silvered glass plate G, and each goes round the axle three times. The two paths are shown exactly superimposed. The four mirrors A, B, C, and D are, of course, fixed.

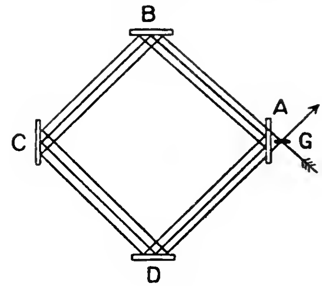


FIG. 327.

If the ether between the discs was dragged with them, it was thought that there might be a displacement of the bands when the motion started. No such displacement was observed. This showed that the ether was not appreciably affected by the moving matter in its vicinity, a result which is in full agreement with Fresnel's formula.

Brace's Experiment. Lord Rayleigh suggested, that if transparent media really contracted in the way described by Fitzgerald and Lorentz when they were moving through space, then they might become doubly-refracting, just as a block of glass does when it is subjected to unilateral stress. He made an attempt to detect such an effect experimentally but with negative results. This experiment was repeated by Brace in 1904 on a more elaborate scale.

Brace used a trough 413 cms. long which was filled with water, but the rays were reflected back and forward along it so that the total length of the path in water was about 30 metres. The trough was rotated about a vertical axis. The light was plane polarised at an angle of 45° to the direction of the earth's motion before it entered the water. If the trough was at right angles to the direction of the earth's motion, and the vibration was resolved into two components perpendicular and parallel to this direction, and if the water contracted in one of these directions, the two components of the vibration might presumably have different velocities. Hence the water might become doubly-refracting. But when the trough was rotated through 90° so that it became parallel to the direction of the earth's motion, the velocity of the two component vibrations should become the same and the double refraction disappear.

No double refraction was observed. Brace estimated that a change in the index of refraction of $7{\cdot}8 \times 10^{-13}$ could have been detected. The greatest effect which might have been expected was 1600 times this.

Lorentz's Theory. While the last experiments described above were performed, the views of physicists as to the nature of light had changed, and they were no longer thinking in terms of the elastic solid theory, but in terms of the electromagnetic theory. The influence of the earth's motion on the electromagnetic theory was investigated very successfully by H. A. Lorentz in 1895.

He starts out from Maxwell's equations in their usual form, referred to axes fixed in space, and assumes that the ions and electrons which form the earth's constitution are all streaming past with velocity v, i.e. the earth's orbital velocity in space. Owing to their rapid motion the electrons are appreciably acted on by the magnetic intensity as well as by the electric intensity of the light wave, and consequently new terms appear in their equations of motion.

He finds as a result that in the moving body the electric intensity can no longer be propagated in plane transversal waves, although the magnetic intensity can. He also finds that V, the velocity of light in a moving body, measured with reference to that moving body in a direction in which the component of the body's velocity is v, is given by

$$\frac{\mathsf{V}}{\mu} - \left(1 - \frac{1}{\mu^2}\right)v,$$

an expression which is the same as Fresnel obtained, except for a slight difference in the definition of μ. Thus the body acts on the wave as if it were dragging with it the medium in which the latter travels, but this drift is purely a virtual one. According to H. A. Lorentz the ether is not a medium in any way analogous to matter, but purely space endowed with the property of propagating wave motion.

Lorentz also finds that the results for a system moving with velocity v in the x direction simplify considerably, when they are expressed as a function of $t - (vx)/\mathsf{V}^2$ instead of as a function of t. Maxwell's equations have then the same form as for a system at rest. Hence $t - (vx)/\mathsf{V}^2$ may be regarded as a sort of local time for the plane specified by x.

The Principle of Relativity. The introduction of this local time and the Fitzgerald-Lorentz contraction have paved the way for a very bold and far-reaching hypothesis, which on account of its difficulty can only be very inadequately mentioned here, but which can hardly be left out on account of its very great importance. This is the principle of relativity introduced by Einstein in 1905, which states that motion through the ether is quite without influence upon all optical experiments made with terrestrial sources of light, and that the apparent velocity of light is a constant.

Suppose that a wave is sent out from a point O on the earth's surface, and that, when this point has reached O_1, the wave-front is a sphere A which has its centre at O. Einstein states that to an observer moving with the point, the wave-front appears to be a sphere with the point O_1 as centre, because events, which appear to be simultaneous to an observer at rest, do not appear to be simultaneous to an observer moving with the system, and because points in the system, which appear to be equidistant to an observer at rest, do not appear to be equidistant to an observer moving with the system.

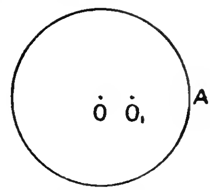


FIG. 328.

The meaning of this sentence may be made clearer by approaching the matter from another side. Suppose a book is lying on the table; then we receive certain sensations of form, colour, touch, smell, etc., from it. We know that these sensations are relative because the colour, for example, may appear different to one who is colour blind, but we attempt to allow for the limitations of our senses and form a concept of the thing itself. Now hitherto, although we assumed that the properties of the book were relative, we always believed that it existed at a certain definite point in time and space, that our time and space were absolute time and space. But according to Einstein they are relative too.

The time and space that an event takes place at depend on the velocity of the coordinate system, and all coordinate systems have the same claim to finality. We cannot get from our relative system to an absolute coordinate system. Now as we have always involuntarily assumed an absolute coordinate system fixed somewhere in the ether, the non-existence of this absolute system appears to many to involve the non-existence of the ether also.

Of course the philosophers have for long written about the relativity of time and space, but their discussions only proved its possibility. The distinctive feature of the new hypothesis is that it is quantitative, and Einstein seeks to show that it is necessary. We have certain phenomena presented to us; according to Einstein we have failed to construct a consistent explanation of them on the assumption that we are dealing with absolute time and space, and hence that assumption must go.

Einstein's principle cuts at the roots of many fundamental ideas, and its applications extend far beyond optics. Hence most people find it exceedingly difficult to get a grip of, and it would be too much to say that it is either generally accepted or denied. But the Michelson-Morley experiment is a somewhat narrow experimental basis on which to rear such a structure.

EXAMPLES.

(1) A plane wave is incident at an angle ϕ on the plane surface of a piece of glass, which is moving with velocity v in the direction of the normal to the surface. Find the value of the angle of reflection by applying Huygens' principle.

(2) Find the angle of refraction in the preceding problem, given that μ is the refractive index of the glass.

TABLES.

TABLE OF RECIPROCALS FOR CALCULATING FOCAL LENGTHS.

Numbers in Difference Columns to be Subtracted, not Added.

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
10	0·0010000	9901	9804	9709	9615	9524	9434	9346	9259	9174	9	18	27	36	45	55	64	73	82
11	0·0009091	9009	8929	8850	8772	8696	8621	8547	8475	8403	8	15	23	30	38	45	53	61	68
12	0·0008333	8264	8197	8130	8065	8000	7937	7874	7813	7752	6	13	19	26	32	38	45	51	58
13	0·0007692	7634	7576	7519	7463	7407	7353	7299	7246	7194	5	11	16	22	27	33	38	44	49
14	0·0007143	7092	7042	6993	6944	6897	6849	6803	6757	6711	5	10	14	19	24	29	33	38	43
15	0·0006667	6623	6579	6536	6494	6452	6410	6369	6329	6289	4	8	13	17	21	25	29	33	38
16	0·0006250	6211	6173	6135	6098	6061	6024	5988	5952	5917	4	7	11	15	18	22	26	29	33
17	0·0005882	5848	5814	5780	5747	5714	5682	5650	5618	5587	3	6	10	13	16	20	23	26	29
18	0·0005556	5525	5495	5464	5435	5405	5376	5348	5319	5291	3	6	9	12	15	17	20	23	26
19	0·0005263	5236	5208	5181	5155	5128	5102	5076	5051	5025	3	5	8	11	13	16	18	21	24
20	0·0005000	4975	4950	4926	4902	4878	4854	4831	4808	4785	2	5	7	10	12	14	17	19	21
21	0·0004762	4739	4717	4695	4673	4651	4630	4608	4587	4566	2	4	7	9	11	13	15	17	20
22	0·0004545	4525	4505	4484	4464	4444	4425	4405	4386	4367	2	4	6	8	10	12	14	16	18
23	0·0004348	4329	4310	4292	4274	4255	4237	4219	4202	4184	2	4	5	7	9	11	13	14	16
24	0·0004167	4149	4132	4115	4098	4082	4065	4049	4032	4016	2	3	5	7	8	10	12	13	15
25	0·0004000	3984	3968	3953	3937	3922	3906	3891	3876	3861	2	3	5	6	8	9	11	12	14
26	0·0003846	3831	3817	3802	3788	3774	3759	3745	3731	3717	1	3	4	6	7	8	10	11	13
27	0·0003704	3690	3676	3663	3650	3636	3623	3610	3597	3584	1	3	4	5	7	8	9	11	12
28	0·0003571	3559	3546	3534	3521	3509	3497	3484	3472	3460	1	2	4	5	6	7	9	10	11
29	0·0003448	3436	3425	3413	3401	3390	3378	3367	3356	3344	1	2	3	5	6	7	8	9	10
30	0·0003333	3322	3311	3300	3289	3279	3268	3257	3247	3236	1	2	3	4	5	6	7	9	10
31	0·0003226	3215	3205	3195	3185	3175	3165	3155	3145	3135	1	2	3	4	5	6	7	8	9
32	0·0003125	3115	3106	3096	3086	3077	3067	3058	3049	3040	1	2	3	4	5	6	7	8	9
33	0·0003030	3021	3012	3003	2994	2985	2976	2967	2959	2950	1	2	3	4	4	5	6	7	8
34	0·0002941	2933	2924	2915	2907	2899	2890	2882	2874	2865	1	2	3	3	4	5	6	7	8
35	0·0002857	2849	2841	2833	2825	2817	2809	2801	2793	2786	1	2	2	3	4	5	6	6	7
36	0·0002778	2770	2762	2755	2747	2740	2732	2725	2717	2710	1	2	2	3	4	5	5	6	7
37	0·0002703	2695	2688	2681	2674	2667	2660	2653	2646	2639	1	1	2	3	4	4	5	6	6
38	0·0002632	2625	2618	2611	2604	2597	2591	2584	2577	2571	1	1	2	3	3	4	5	5	6
39	0·0002564	2558	2551	2545	2538	2532	2525	2519	2513	2506	1	1	2	3	3	4	4	5	6
40	0·0002500	2494	2488	2481	2475	2469	2463	2457	2451	2445	1	1	2	2	3	4	4	5	5
41	0·0002439	2433	2427	2421	2415	2410	2404	2398	2392	2387	1	1	2	2	3	3	4	5	5
42	0·0002381	2375	2370	2364	2358	2353	2347	2342	2336	2331	1	1	2	2	3	3	4	4	5
43	0·0002326	2320	2315	2309	2304	2299	2294	2288	2283	2278	1	1	2	2	3	3	4	4	5
44	0·0002273	2268	2262	2257	2252	2247	2242	2237	2232	2227	1	1	2	2	3	3	4	4	5
45	0·0002222	2217	2212	2208	2203	2198	2193	2188	2183	2179	0	1	1	2	2	3	3	4	4
46	0·0002174	2169	2165	2160	2155	2151	2146	2141	2137	2132	0	1	1	2	2	3	3	4	4
47	0·0002128	2123	2119	2114	2110	2105	2101	2096	2092	2088	0	1	1	2	2	3	3	4	4
48	0·0002083	2079	2075	2070	2066	2062	2058	2053	2049	2045	0	1	1	2	2	3	3	3	4
49	0·0002041	2037	2033	2028	2024	2020	2016	2012	2008	2004	0	1	1	2	2	2	3	3	4
50	0·0002000	1996	1992	1988	1984	1980	1976	1972	1969	1965	0	1	1	2	2	2	3	3	4
51	0·0001961	1957	1953	1949	1946	1942	1938	1934	1931	1927	0	1	1	2	2	2	3	3	3
52	0·0001923	1919	1916	1912	1908	1905	1901	1898	1894	1890	0	1	1	1	2	2	3	3	3
53	0·0001887	1883	1880	1876	1873	1869	1866	1862	1859	1855	0	1	1	1	2	2	2	3	3
54	0·0001852	1848	1845	1842	1838	1835	1832	1828	1825	1821	0	1	1	1	2	2	2	3	3

TABLE OF RECIPROCALS FOR CALCULATING FOCAL LENGTHS.

Numbers in Difference Columns to be Subtracted, not Added.

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
55	0·0001818	1815	1812	1808	1805	1802	1799	1795	1792	1789	0	1	1	1	2	2	2	3	3
56	0·0001786	1783	1779	1776	1773	1770	1767	1764	1761	1757	0	1	1	1	2	2	2	3	3
57	0·0001754	1751	1748	1745	1742	1739	1736	1733	1730	1727	0	1	1	1	2	2	2	2	3
58	0·0001724	1721	1718	1715	1712	1709	1706	1704	1701	1698	0	1	1	1	1	2	2	2	3
59	0·0001695	1692	1689	1686	1684	1681	1678	1675	1672	1669	0	1	1	1	1	2	2	2	3
60	0·0001667	1664	1661	1658	1656	1653	1650	1647	1645	1642	0	1	1	1	1	2	2	2	3
61	0·0001639	1637	1634	1631	1629	1626	1623	1621	1618	1616	0	1	1	1	1	2	2	2	2
62	0·0001613	1610	1608	1605	1603	1600	1597	1595	1592	1590	0	1	1	1	1	2	2	2	2
63	0·0001587	1585	1582	1580	1577	1575	1572	1570	1567	1565	0	0	1	1	1	1	2	2	2
64	0·0001563	1560	1558	1555	1553	1550	1548	1546	1543	1541	0	0	1	1	1	1	2	2	2
65	0·0001538	1536	1534	1531	1529	1527	1524	1522	1520	1517	0	0	1	1	1	1	2	2	2
66	0·0001515	1513	1511	1508	1506	1504	1502	1499	1497	1495	0	0	1	1	1	1	2	2	2
67	0·0001493	1490	1488	1486	1484	1481	1479	1477	1475	1473	0	0	1	1	1	1	2	2	2
68	0·0001471	1468	1466	1464	1462	1460	1458	1456	1453	1451	0	0	1	1	1	1	2	2	2
69	0·0001449	1447	1445	1443	1441	1439	1437	1435	1433	1431	0	0	1	1	1	1	2	2	2
70	0·0001429	1427	1425	1422	1420	1418	1416	1414	1412	1410	0	0	1	1	1	1	1	2	2
71	0·0001408	1406	1404	1403	1401	1399	1397	1395	1393	1391	0	0	1	1	1	1	1	2	2
72	0·0001389	1387	1385	1383	1381	1379	1377	1376	1374	1372	0	0	1	1	1	1	1	2	2
73	0·0001370	1368	1366	1364	1362	1361	1359	1357	1355	1353	0	0	1	1	1	1	1	2	2
74	0·0001351	1350	1348	1346	1344	1342	1340	1339	1337	1335	0	0	1	1	1	1	1	1	2
75	0·0001333	1332	1330	1328	1326	1325	1323	1321	1319	1318	0	0	1	1	1	1	1	1	2
76	0·0001316	1314	1312	1311	1309	1307	1305	1304	1302	1300	0	0	1	1	1	1	1	1	2
77	0·0001299	1297	1295	1294	1292	1290	1289	1287	1285	1284	0	0	0	1	1	1	1	1	1
78	0·0001282	1280	1279	1277	1276	1274	1272	1271	1269	1267	0	0	0	1	1	1	1	1	1
79	0·0001266	1264	1263	1261	1259	1258	1256	1255	1253	1252	0	0	0	1	1	1	1	1	1
80	0·0001250	1248	1247	1245	1244	1242	1241	1239	1238	1236	0	0	0	1	1	1	1	1	1
81	0·0001235	1233	1232	1230	1229	1227	1225	1224	1222	1221	0	0	0	1	1	1	1	1	1
82	0·0001220	1218	1217	1215	1214	1212	1211	1209	1208	1206	0	0	0	1	1	1	1	1	1
83	0·0001205	1203	1202	1200	1199	1198	1196	1195	1193	1192	0	0	0	1	1	1	1	1	1
84	0·0001190	1189	1188	1186	1185	1183	1182	1181	1179	1178	0	0	0	1	1	1	1	1	1
85	0·0001176	1175	1174	1172	1171	1170	1168	1167	1166	1164	0	0	0	1	1	1	1	1	1
86	0·0001163	1161	1160	1159	1157	1156	1155	1153	1152	1151	0	0	0	1	1	1	1	1	1
87	0·0001149	1148	1147	1145	1144	1143	1142	1140	1139	1138	0	0	0	1	1	1	1	1	1
88	0·0001136	1135	1134	1133	1131	1130	1129	1127	1126	1125	0	0	0	1	1	1	1	1	1
89	0·0001124	1122	1121	1120	1119	1117	1116	1115	1114	1112	0	0	0	1	1	1	1	1	1
90	0·0001111	1110	1109	1107	1106	1105	1104	1103	1101	1100	0	0	0	1	1	1	1	1	1
91	0·0001099	1098	1096	1095	1094	1093	1092	1091	1089	1088	0	0	0	0	1	1	1	1	1
92	0·0001087	1086	1085	1083	1082	1081	1080	1079	1078	1076	0	0	0	0	1	1	1	1	1
93	0·0001075	1074	1073	1072	1071	1070	1068	1067	1066	1065	0	0	0	0	1	1	1	1	1
94	0·0001064	1063	1062	1060	1059	1058	1057	1056	1055	1054	0	0	0	0	1	1	1	1	1
95	0·0001053	1052	1050	1049	1048	1047	1046	1045	1044	1043	0	0	0	0	1	1	1	1	1
96	0·0001042	1041	1040	1038	1037	1036	1035	1034	1033	1032	0	0	0	0	1	1	1	1	1
97	0·0001031	1030	1029	1028	1027	1026	1025	1024	1022	1021	0	0	0	0	1	1	1	1	1
98	0·0001020	1019	1018	1017	1016	1015	1014	1013	1012	1011	0	0	0	0	1	1	1	1	1
99	0·0001010	1009	1008	1007	1006	1005	1004	1003	1002	1001	0	0	0	0	0	1	1	1	1

Indices of Refraction of Optical Glasses made by Chance Brothers & Co., Limited, Birmingham.

Factory Number.	Name.	μ_D.	$\mu_D - \mu_C$.	$\mu_F - \mu_D$.	$\mu_G - \mu_F$.
646	Boro-silicate crown	1·5096	0·00236	0·00562	0·00446
605	Hard crown	1·5175	0·00252	0·00604	0·00484
577	Medium barium crown	1·5738	0·00293	0·00697	0·00552
579	Densest barium crown	1·6065	0·00308	0·00738	0·00589
569	Soft crown	1·5152	0·00264	0·00642	0·00517
563	Medium barium crown	1·5660	0·00297	0·00709	0·00576
535	Barium light flint	1·5452	0·00298	0·00722	0·00582
490	Extra light flint	1·5316	0·00313	0·00772	0·00630
485	Extra light flint	1·5333	0·00322	0·00777	0·00640
466	Barium light flint	1·5833	0·00362	0·00889	0·00721
458	Soda flint	1·5482	0·00343	0·00852	0·00690
458	Light flint	1·5472	0·00348	0·00848	0·00707
432	Light flint	1·5610	0·00372	0·00927	0·00770
410	Light flint	1·5760	0·00402	0·01002	0·00840
407	Light flint	1·5787	0·00404	0·01016	0·00840
370	Dense flint	1·6118	0·00470	0·01187	0·01004
361	Dense flint	1·6214	0·00491	0·01231	0·01046
360	Dense flint	1·6225	0·00493	0·01236	0·01054
337	Extra dense flint	1·6469	0·00541	0·01376	0·01170
299	Double extra dense flint	1·7129	0·00670	0·01714	0·01661

Refraction and Dispersion of Gases.

The second column gives the index of refraction for sodium light reduced to a standard density at 0° C. and 760 mm. on the assumption that $(\mu - 1)/\rho$ is constant. The third and fourth columns give the constants of the formula $\mu - 1 = A(1 + B/\lambda^2)$, where λ is measured in cms.; this formula represents the index of refraction throughout the visible spectrum.

Gas.	μ_D.	A.	B.
Air	1·0002918	$28·71\ 10^{-5}$	$5·67\ 10^{-11}$
Hydrogen	1·0001384	13·58 „	7·52 „
Helium	1·0000350	3·48 „	2·3 „
Oxygen	1·000272	26·63 „	5·07 „
Nitrogen	1·000297	29·06 „	7·7 „

INDICES OF REFRACTION OF SOLIDS AND LIQUIDS WITH RESPECT TO AIR.

Wave-length.	Water at 20° C.	Carbon Bisulphide.		Quartz at 18° C.		Iceland Spar at 18° C.		Fluorite at 18°.	Rock Salt at 18°.
		At 0°.	At 20°.	Ord.	Ext.	Ord.	Ext.		
Ultra-violet									
Al 1857	—	—	—	1·6750	1·6891	—	—	1·5094	1·8933
Cd 2144	1·4040	—	—	1·6304	1·6427	1·8456	1·5598	1·4848	1·7322
Cd 2313	1·3888	—	—	1·6140	1·6256	1·8023	1·5454	1·4752	1·6884
Cd 2749	1·3664	2·0348	2·0047	1·5875	1·5981	1·7414	1·5226	1·4596	1·6269
Cd 3404	1·3504	—	—	1·5674	1·5774	1·7008	1·5056	1·4478	1·5860
Cd 3612	1·3474	1·7572	1·7381	1·5634	1·5732	1·6932	1·5022	1·4453	1·5784
Visible									
H 3968	1·3435	1·7199	1·7018	1·5581	1·5677	1·6832	1·4977	1·4421	1·5682
H 4340	1·3404	—	—	1·5540	1·5634	1·6755	1·4943	1·4396	1·5612
H 4861	1·3371	1·6713	1·6547	1·5497	1·5590	1·6678	1·4907	1·4371	1·5534
Na 5893	1·3330	1·6436	1·6276	1·5442	1 5533	1·6584	1·4864	1·4338	1·5443
Li 6708	1·3307	1·6328	1·6168	—	—	1·6537	1·4843	1·4323	1·5400
K 7685	1·3289	1·6241	1·6087	1·5390	1·5479	1·6497	1·4826	1·4309	1·5367
Infra-red									
1·0 μ	1·3249	—	1·5968	1·5350	1·5437	1·6436	1·4801	1·4290	1·5321
2·0 ,,	—	—	1·584	1·521	1·529	1·625	1·475	1·4239	1·5267
3·0 ,,	—	—	—	1·499	—	—	—	1·4179	1·5241
4·0 ,,	—	—	—	1·465	—	—	—	1·4097	1·5218

Carbon bisulphide is used frequently as a standard substance in determining refractive indices by total reflection methods. Its refractive index varies rapidly with the temperature. Hence it is given both for 0° and 20° C.

INDICES OF REFRACTION FOR SODIUM LIGHT AT 15° C.

Alcohol, amyl	1·41	Ether, ethyl	1·352
Alcohol, ethyl	1·362	Glycerine	1·463
Alcohol, methyl	1·332	Ice	1·31
Aniline	1·590	Mica	1·56 to 1·60
Benzene	1·504	Methylene iodide	1·742
Canada balsam	1·53	Monobrom naphthalene	1·660
Cedar oil	1·516	Paraffin oil	1·44
Chloroform	1·446	Ruby	1·76
Diamond	2·417	Sugar	1·54 to 1·57
		Turpentine	1·48

ROTATION IN QUARTZ PER MM. AT 20° C.

H 4102.	H 4861.	Tl 5351.	Na 5893.	H 6563.	Li 6708.
47·48°	32·7	26·53	21·72	17·3	16·4

Percentage of Light Reflected by Metal Mirrors. (E. Hagen and H. Rubens.)

Wave-length.	Speculum Metal. 68·2 Cu + 31·8 Sn.	Electrolytically Deposited Nickel.	Electrolytically Deposited Platinum.	Chemically Deposited Silver.
Ultra-violet				
·251 μ	29·9	37·8	33·8	34·1
·288	37·7	42·7	38·8	21·2
·305	41·7	44·2	39·8	9·1
·316	—	—	—	4·2
·326	—	45·2	41·4	14·6
·338	—	46·5	—	55·5
·357	51·0	48·8	43·4	74·5
·385	53·1	49·6	45·4	81·4
Visible				
·420	56·4	56·6	51·8	86·6
·450	60·0	59·4	54·7	90·5
·500	63·2	60·8	58·4	91·3
·550	64·0	62·6	61·1	92·7
·600	64·3	64·9	64·2	92·6
·650	65·4	66·6	66·5	93·5
·700	66·8	68·8	69·0	94·6
Infra-red				
·800	—	69·6	70·3	96·3
1·00	70·5	72·0	72·9	96·6
1·5	75·0	78·6	77·7	98·4
2·0	80·4	83·5	80·6	
3·0	86·2	88·7	88·8	
4·0	88·5	91·1	91·5	
5·0	89·1	94·4	93·5	
7·0	90·1	94·3	95·5	
9·0	92·2	95·6	95·4	
11·0	92·9	95·9	95·6	
14·0	93·6	97·2	96·4	

Drude's Values for the Optical Constants of Metals for Na Light.

Metal.	κ.	ν.
Silver	3·67	0·18
Gold	2·82	0·37
Platinum	4·26	2·06
Copper	2·62	0·64
Steel	3·40	2·41
Sodium	2·61	0·005
Mercury	4·96	1·73

Calibration of the Ultra-violet.

The following spectral lines are useful for calibrating a spectrograph of moderate dispersion. They are produced by passing the spark in air between terminals of the metals in question, a condenser being connected up in parallel with the spark. The zinc lines printed in heavy type are prominent and easily recognised. The air lines become stronger if the condenser is removed. The three aluminium lines require a very long exposure. The sensitiveness of the photographic plate is low for wave-lengths smaller than 2300. The relative brightness of the different lines depends on the induction coil.

Al 1852·2	Cd 2194·7	Zn 3072·2	Pb 4060	Air 5006·0
,, 1933·5	,, 2312·9	,, 3282·4	,, 4246·6	,, 5679·1
,, 1988·1	Zn **2516·0**	,, **3303·0**	,, 4387·3	,, 5933·1
Zn **2024·3**	,, **2558·0**	,, **3345·1**	Zn **4722·1**	,, 5942·6
,, **2061·0**	,, 2771·0	Cd 3404	,, **4810·5**	Zn 6103·0
,, **2098·8**	,, 2801·0	,, 3467	,, **4912·0**	,, 6363·7
,, **2138·3**	,, 3035·9	,, 3612	Air 5003·0	

Wave-length and Colour of the Principal Fraunhofer Lines.

The first column gives the name of the line, the second the element that produces it, and the third the wave-length in A.U.

Indigo and Violet, 3600-4550		
K	Ca	3933·6
H	Ca	3968·4
h	H	4101·8
g	Ca	4226·7
G	Fe, Ca *	4307·9
G^1	Fe	4325·8
f	H	4340·4
e	Fe	4383·6
Blue, 4550-4920		
d	Fe	4668
F	H	4861·4
Green, 4920-5500		
c	Fe	4957·6
b_4	Mg, Fe *	5167·4
b_3	Fe	5169·0
b_2	Mg	5172·7
b_1	Mg	5178·2
E	Fe	5269·6
Yellow, 5500-5880		
Orange, 5880-6470		
D_3	He	5875·6
D_2	Na	5890·0
D_1	Na	5895·9
a	O	6278·1
Red, 6470-7700		
C	H	6562·8
B	O	6870
a		7185
A		7661

* There are two lines here, one belonging to each element, which can be separated by a high resolving power.

INDEX.

The references are to pages.

PRINTED IN GREAT BRITAIN BY THE UNIVERSITY PRESS, ABERDEEN